

Breaking Gender Barriers:
Experimental Evidence on Men in Pink-Collar Jobs

Online Appendix

Alexia Delfino*

*Bocconi University, alexia.delfino@unibocconi.it.

Contents

- A Appendix figures and tables 3

- B Methodology for the creation of indexes 34
 - B.1 Index of desirable skills 34
 - B.2 Predicted outside option: wage level and dispersion 34
 - B.3 Predicted performance on the job 35
 - B.4 Index of occupational gender segregation 35

- C Out-of-trial surveys: implementation 36

- D External validity of the photographs effect: a complementary field experiment 39

- E Information interpretation: evidence from a survey with open-ended questions 40

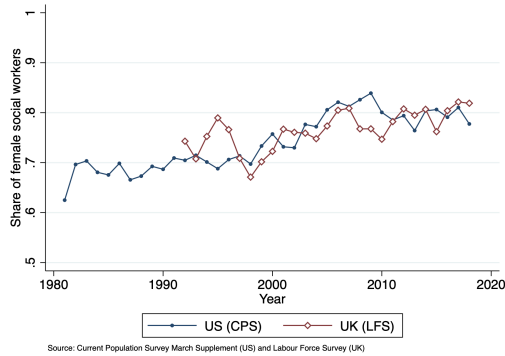
- F Additional alternative treatment interpretations 41

- References 42

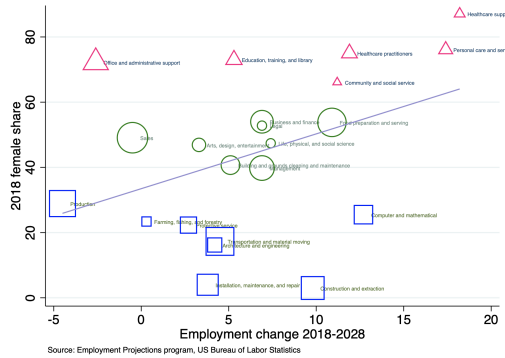
A Appendix figures and tables

Figure A.1. Three facts about social work

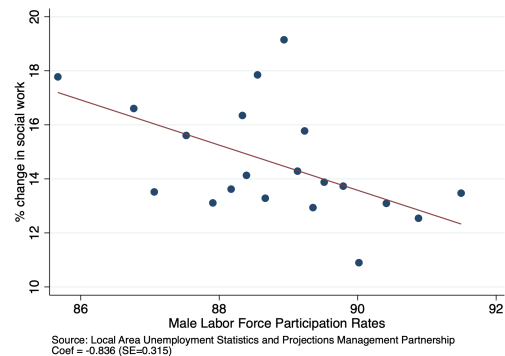
(a) Historical female shares in social work



(b) Predicted growth in occupations by workers' female share (2018-2028)



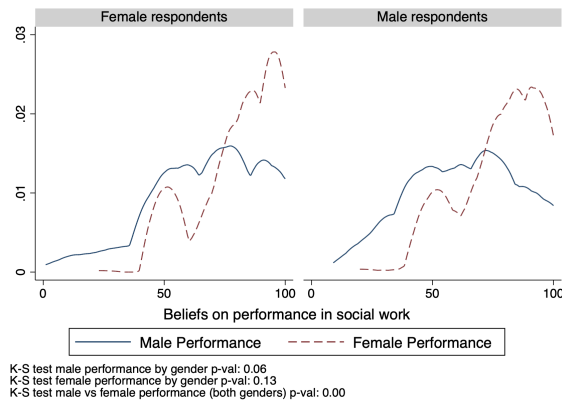
(c) Social work predicted growth and male labor force participation



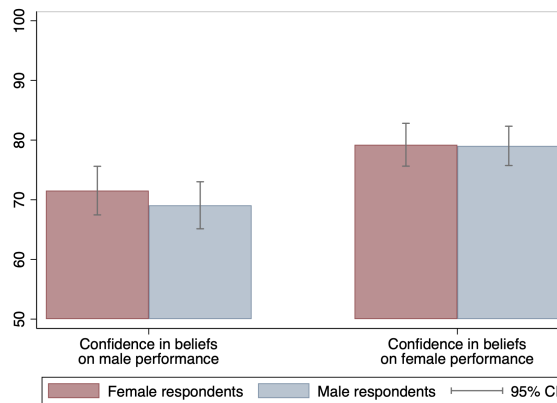
Note. Panel (a) shows the female share of social workers from 1980 to 2018 in the US and UK. Data are from the Current Population Survey (Flood et al., 2023) and the Labour Force Survey (Office for National Statistics, 2023). Panel (b) shows the correlation between the predicted percentage change in employment between 2018 and 2028 (on the x-axis) and the 2018 female share of workers (on the y-axis) for major occupational groups in the US. Data are from the US Bureau of Labor Statistics (2018b) and the Employment Projections program (US Bureau of Labor Statistics, 2019). Panel (c) shows a binned scatterplot between the 2018 male labor force participation (on the x-axis) and the percentage change in employment in social work between 2018 and 2028 (on the y-axis) across US states. The graph controls for the overall growth rate across occupations and the state-level female labor force participation. Data are from the Local Area Unemployment Statistics (US Bureau of Labor Statistics, 2017) and Occupational Projections by the US Department of Labor (Projections Central, 2016).

Figure A.2. Beliefs about workers' performance and returns to talent in social work

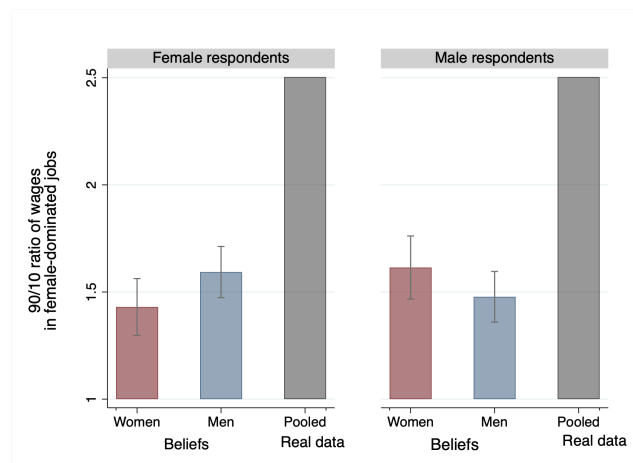
(a) Beliefs on workers' performance



(b) Confidence in beliefs on workers' performance



(c) Expected returns to talent



Note. Panel (a) shows densities of answers to the question: “On a scale from 0 to 100, what do you think is the performance of a [WOMAN/MAN] in social work?”. The graph on the LHS (RHS) shows the distribution of women’s (men’s) beliefs. Dashed lines represent beliefs about women’s performance and solid blue lines about men’s performance. Panel (b) shows answers to: “On a scale from 0 to 100, how confident are you of your answer [about the performance of a man/woman in social work]?”. Dark blue bars show answers for female respondents and lighter red bars for male respondents. Panel (c) shows the average beliefs on the ratio of wages between percentiles 90 and 10 in three female-dominated jobs (social work, nursing and elementary school teaching), by respondents’ gender and by workers’ gender. The last bar shows the average ratio for the same occupations from labor market data from the UK Annual Survey of Hours and Earnings ([Office for National Statistics, 2020](#)) and, when not available, for the US ([US Bureau of Labor Statistics, 2018a](#)). Data for (a) and (b) come from the auxiliary surveys discussed in Section V. Data for (c) come from an additional survey collected on Prolific (“Survey 5” described in Appendix C).

Table A.1. Out-of-trial surveys: checks on workers’ photographs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Female Photo			Male Photo			Diff means
	Mean	SD	N	Mean	SD	N	P-val
Panel A: 2018 Applicants							
<i>White pictures</i>							
Friendliness	0.79	0.41	92	0.63	0.49	94	0.01
Work satisfaction	0.91	0.28	92	0.84	0.37	94	0.13
<i>Non-white pictures</i>							
Friendliness	0.86	0.36	28	0.82	0.39	28	0.72
Work satisfaction	0.82	0.39	28	0.93	0.26	28	0.23
Panel B: Prolific sample							
<i>White pictures</i>							
Friendliness	0.87	0.34	98	0.74	0.44	95	0.02
Work satisfaction	0.81	0.4	98	0.76	0.43	95	0.42
Happy feeling	0.57	0.5	98	0.31	0.46	95	0
<i>Non-white pictures</i>							
Friendliness	0.97	0.17	33	0.92	0.28	36	0.35
Work satisfaction	0.97	0.17	33	0.92	0.28	36	0.35
Happy feeling	0.76	0.44	33	0.67	0.48	36	0.41
Panel C: Amazon Turk sample							
<i>White pictures</i>							
Friendliness	0.9	0.31	39	0.74	0.45	38	0.07
Work satisfaction	0.87	0.34	39	0.76	0.43	38	0.22
Happy feeling	0.79	0.41	39	0.66	0.48	38	0.18
Trust	0.85	0.37	39	0.82	0.39	38	0.73
Attractiveness	0.72	0.46	39	0.76	0.43	38	0.66
Professional clothing	0.38	0.49	39	0.87	0.34	38	0.00
<i>Non-white pictures</i>							
Friendliness	0.98	0.15	42	0.95	0.22	42	0.56
Work satisfaction	0.95	0.22	42	0.88	0.33	42	0.24
Happy feeling	0.9	0.3	42	0.9	0.3	42	1.00
Trust	0.93	0.26	42	0.88	0.33	42	0.46
Attractiveness	0.95	0.22	42	0.74	0.45	42	0.01
Professional clothing	0.93	0.26	42	0.9	0.3	42	0.70

Note. The table shows survey respondents’ opinions about the photographs used in the experiment. Each respondent was shown only one photograph, and had no additional information about its intended use. “Friendliness” and “Work satisfaction” are indicators equal to one if the respondent rated the photographed worker as friendly or very friendly, and satisfied or very satisfied with his/her work. The variable “Happy feeling” takes values between -3 (“Extremely unhappy”) and 3 (“Extremely happy”), and it is constructed from the question “To what extent does this image make you feel happy?”. “Trust” is defined from the question “If this person was giving you some information about her job, would you trust him/her?”, and has values between -2 (“Definitely not”) and 2 (“Definitely yes”). “Attractiveness” rates the photograph between -2 (“Not attractive”) and 2 (“Attractive”). “Professional clothing” is a dummy equal to one if the respondent would describe the clothes of the portrayed person as “professional” (0 if “unprofessional”). Data for Panels A and B come from the out-of-trial surveys discussed in Section V. Data for Panel C come from an additional survey conducted with Amazon Turk workers (“Survey 1” described in Appendix C).

Table A.2. Treatment effect on applications by ethnicity

	(1)	(2)
	Applied & no drop-out	
	White	Non-White
Panel A: Men		
Male Photo	0.012 (0.041) [0.78]	-0.090 (0.066) [0.17]
Lower Share	0.055 (0.041) [0.18]	0.114 (0.065) [0.07]
Photo=Info p	0.47	0.03
Mean DV	0.49	0.59
Mean DV in PC	0.55	0.48
N	586	221
Male Photo: White=Non-White p		0.18
Lower Share: White=Non-White p		0.44
Panel B: Women		
Male Photo	-0.050 (0.019) [0.01]	-0.054 (0.031) [0.08]
Lower Share	-0.016 (0.019) [0.40]	-0.013 (0.031) [0.67]
Photo=Info p	0.21	0.35
Mean DV	0.59	0.62
Mean DV in PC	0.58	0.61
N	2555	958
Male Photo: White=Non-White p		0.93
Lower Share: White=Non-White p		0.93
<i>Gender differences</i>		
Photo: Men = Women p	0.17	0.61
Info: Men = Women p	0.12	0.08

Robust standard errors in parentheses

Randomization Inference p-values in square brackets

Note. The table shows OLS estimates controlling for access to early registration (stratification variable). Each panel reports results of two different regressions splitting the sample by candidates of white (Column 1) or non-white (Column 2) ethnicity. The dependent variable is an indicator for applying and never dropping out of the process. The rows “Male Photo: White=Non-White p” and “Lower Share: White=Non-White p” report the p-value of a test of equality of the coefficients on the “Male Photo” and “Lower Share,” respectively, for the white versus non-white samples. For an additional description of this table, see the notes for Table 2.

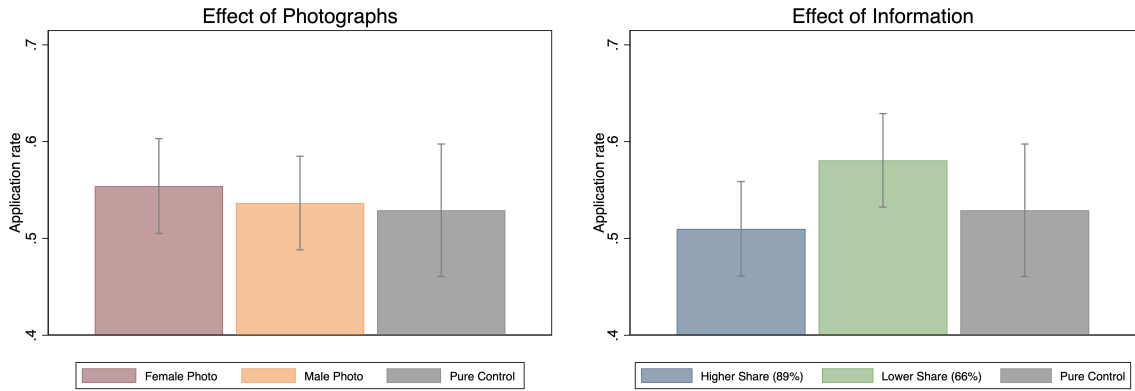
Table A.3. Comparison of the experimental sample with UK students and workers

Sample:	(1)	(2)	(3)	(4)	(5)	(6)
	Men			Women		
	Experiment	UK	Diff (1)-(2)	Experiment	UK	Diff (1)-(2)
Panel A: Higher Education Statistics Agency (HESA)						
Engineering	0.02	0.13	-0.12	0.01	0.02	-0.02
Math & computing	0.02	0.12	-0.10	0.01	0.03	-0.02
Business & finance	0.08	0.17	-0.09	0.05	0.13	-0.08
Medicine, dentistry & allied	0.02	0.09	-0.07	0.03	0.2	-0.17
Physical sciences	0.01	0.06	-0.04	0.01	0.03	-0.02
Architecture	0.00	0.03	-0.03	0.00	0.01	-0.01
Arts	0.05	0.06	-0.02	0.04	0.09	-0.05
Agriculture	0.00	0.01	-0.01	0.00	0.01	-0.01
Mass communication	0.02	0.02	0.00	0.01	0.02	-0.01
Education	0.04	0.03	0.01	0.05	0.08	-0.03
Law	0.04	0.03	0.01	0.04	0.04	0.00
Biological sciences	0.13	0.08	0.04	0.21	0.11	0.10
Languages	0.08	0.03	0.05	0.09	0.06	0.04
Humanities	0.15	0.04	0.11	0.07	0.04	0.04
Social studies	0.34	0.08	0.26	0.38	0.11	0.27
Panel B: Labour Force Survey (LFS)						
Age	28.68	29.3	-0.63	26.35	28.77	-2.43
Graduate	0.46	0.75	-0.29	0.35	0.73	-0.39
FTE health	0.15	0.18	-0.03	0.16	0.41	-0.25
Married	0.19	0.27	-0.08	0.12	0.28	-0.16
First grade	0.20	0.14	0.06	0.18	0.15	0.04
FTE public sector	0.45	0.27	0.18	0.55	0.49	0.06
Non-white ethnicity	0.28	0.14	0.14	0.27	0.12	0.15

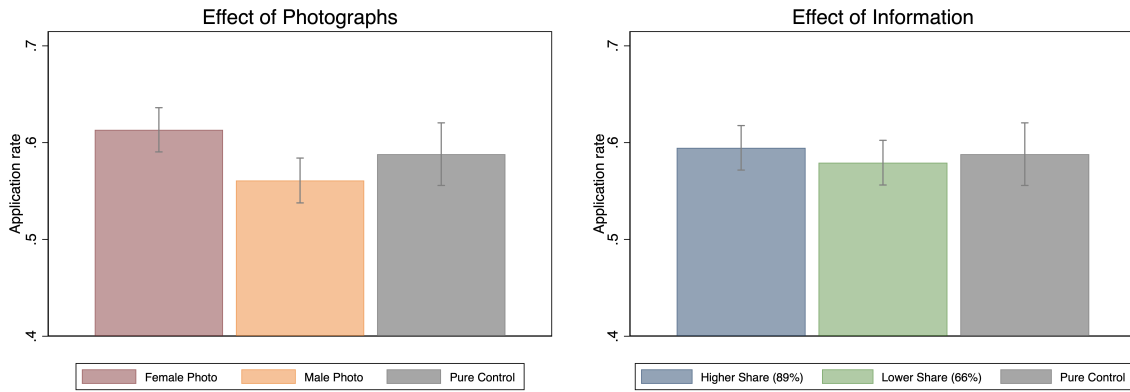
Note. The table compares educational and demographic variables for the men and women in the experimental sample (Columns (1) and (4)) with UK students and workers (Columns (2) and (5)). Panel A shows the distribution of male and female university students across 15 subject categories, with Columns (3) and (6) showing the difference in shares between the experimental sample and UK averages. Data are from the Higher Education Statistics Agency for the academic year 2017-2018 ([Higher Education Statistics Agency, 2017a](#)). Panel B uses data from a random sample of the 2017 and 2018 Labour Force Survey ([Office for National Statistics, 2023](#)) generated to reproduce the same age distribution of the experimental sample. “FTE public sector” is an indicator variable for working in the government, “FTE health” is an indicator variable for working in healthcare. “First grade” is a dummy for achieving the highest grade in university. “Graduate” indicates whether the candidate had already graduated in the year before the RCT.

Figure A.3. Application rates by treatment and gender

(a) Men



(b) Women



Note. Panel A shows application rates for men by photograph treatment (left-hand side) and information treatment (right-hand side). Panel B shows application rates for women by photograph treatment (left-hand side) and information treatment (right-hand side). Error bars show 95% confidence intervals. The application outcome variable is defined in Section III.A.

Table A.4. Application outcomes: controls for strata only

	(1)	(2)	(3)
	Application outcomes		
	Applied & no drop-out	Received offer	Accepted offer
Panel A: Men			
Male Photo	-0.016 (0.035) [0.62] {0.64}	0.030 (0.019) [0.11] {0.22}	0.026 (0.015) [0.09] {0.22}
Lower Share	0.071 (0.035) [0.04] {0.08}	0.042 (0.019) [0.02] {0.07}	0.022 (0.016) [0.17] {0.16}
Photo=Info p	0.08	0.66	0.86
Mean DV	0.52	0.05	0.03
Mean DV in PC	0.53	0.11	0.09
N	807	807	807
Panel B: Women			
Male Photo	-0.051 (0.017) [0.00] {0.01}	0.000 (0.009) [0.97] {0.96}	0.012 (0.008) [0.12] {0.18}
Lower Share	-0.015 (0.017) [0.36] {0.65}	-0.000 (0.009) [0.98] {0.96}	-0.001 (0.008) [0.95] {0.99}
Photo=Info p	0.12	0.95	0.25
Mean DV	0.60	0.09	0.05
Mean DV in PC	0.59	0.09	0.06
N	3513	3513	3513
<i>Gender differences</i>			
Photo: Men=Women p	0.36	0.16	0.41
Info: Men=Women p	0.03	0.04	0.19

Robust standard errors in parentheses
Randomization Inference p-values in square brackets
FWER corrected p-values in curly brackets

Note. The table repeats the estimations shown in Table 2, but controlling only for stratification variables (ethnicity and early registration). See the notes for Table 2 for a detailed description of this table.

Table A.5. Application outcomes: conditional and unconditional variables

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Applied & no drop-out	Applied	Drop-Out		Received offer		Accepted offer	
			Uncond	Cond on App	Uncond	Cond on App	Uncond	Cond on App
Panel A: Men								
Male Photo	-0.023 (0.035) [0.50]	-0.037 (0.035) [0.26]	-0.009 (0.016) [0.56]	-0.014 (0.026) [0.60]	0.033 (0.020) [0.09]	0.064 (0.035) [0.06]	0.028 (0.016) [0.08]	0.052 (0.029) [0.07]
Lower Share	0.075 (0.035) [0.03]	0.053 (0.035) [0.13]	-0.016 (0.016) [0.35]	-0.039 (0.027) [0.15]	0.049 (0.020) [0.01]	0.057 (0.034) [0.09]	0.027 (0.016) [0.11]	0.034 (0.029) [0.25]
Photo=Info p	0.05	0.07	0.79	0.51	0.57	0.89	0.94	0.65
Mean DV	0.52	0.59	0.07	0.13	0.05	0.10	0.03	0.07
Mean DV in PC	0.53	0.58	0.05	0.08	0.11	0.21	0.09	0.17
N	807	807	807	485	807	440	807	440
Panel B: Women								
Male Photo	-0.051 (0.016) [0.00]	-0.052 (0.016) [0.00]	-0.001 (0.007) [0.90]	0.005 (0.011) [0.64]	-0.001 (0.009) [0.94]	0.015 (0.015) [0.34]	0.011 (0.008) [0.14]	0.030 (0.013) [0.03]
Lower Share	-0.014 (0.016) [0.41]	-0.009 (0.016) [0.58]	0.005 (0.007) [0.45]	0.011 (0.011) [0.33]	-0.001 (0.009) [0.90]	0.010 (0.015) [0.54]	-0.000 (0.008) [0.97]	0.004 (0.013) [0.72]
Photo=Info p	0.11	0.06	0.55	0.73	0.96	0.83	0.30	0.16
Mean DV	0.60	0.64	0.04	0.06	0.09	0.14	0.05	0.08
Mean DV in PC	0.59	0.62	0.03	0.05	0.09	0.15	0.06	0.10
N	3513	3513	3513	2229	3513	2062	3513	2062
<i>Gender Differences</i>								
Photo: Men=Women p	0.46	0.70	0.63	0.51	0.12	0.19	0.34	0.47
Info: Men=Women p	0.02	0.10	0.23	0.08	0.02	0.20	0.13	0.34

Robust standard errors in parentheses
Randomization inference p-values in square brackets

Note. The table shows OLS estimates with PDS Lasso controls. Each panel reports the results of eight regressions. Columns (1), (5) and (7) report the same regressions of Table 2. In Column (2), the dependent variable is an indicator for applying in the first stage of the hiring process. The outcome variable in Columns (3) and (4) is an indicator equal to one if a person, after applying in stage one, voluntarily withdraws from the hiring process at any later stage. Column (3) reports results on unconditional withdrawals and Column (4) only considers people who applied in stage one. Column (6) reports results on receiving a job offer, conditional on applying. Column (8) reports results on offer acceptance conditional on applying. For an additional description of this table, see the notes for Table 2.

Table A.6. Application outcomes for men: fully interacted treatments

	(1)	(2)	(3)	(4)	(5)	(6)
	PDS Lasso Controls			DML Controls		
	Applied	Offer	Accepted	Applied	Offer	Accepted
T1: Female Ph + Higher Sh	-0.008 (0.049) [0.87]	-0.065 (0.027) [0.02]	-0.060 (0.024) [0.01]	-0.003 (0.049) [0.55]	-0.063 (0.026) [0.01]	-0.050 (0.024) [0.03]
T2: Female Ph + Lower Sh	0.071 (0.050) [0.14]	-0.021 (0.030) [0.48]	-0.046 (0.026) [0.07]	0.061 (0.050) [0.08]	-0.029 (0.030) [0.41]	-0.046 (0.025) [0.21]
T3: Male Ph + Higher Sh	-0.029 (0.050) [0.53]	-0.036 (0.029) [0.22]	-0.043 (0.026) [0.11]	-0.034 (0.048) [0.71]	-0.041 (0.027) [0.35]	-0.035 (0.024) [0.28]
T4: Male Ph + Lower Sh	0.051 (0.049) [0.29]	0.010 (0.032) [0.76]	-0.012 (0.028) [0.66]	0.058 (0.050) [0.15]	0.011 (0.031) [0.72]	-0.003 (0.027) [0.96]
<i>Overall effects</i>						
Lower Sh vs PC: $\frac{T2+T4}{2}$	0.06 (0.04)	-0.01 (0.03)	-0.03 (0.02)	0.06 (0.04)	-0.01 (0.03)	-0.03 (0.02)
Lower Sh vs Higher Sh: $\frac{(T2+T4)-(T1+T3)}{2}$	0.08 (0.04)	0.04 (0.02)	0.02 (0.02)	0.08 (0.04)	0.04 (0.02)	0.02 (0.02)
Male Ph vs PC: $\frac{T3+T4}{2}$	0.01 (0.04)	-0.01 (0.03)	-0.03 (0.02)	0.01 (0.04)	-0.01 (0.03)	-0.02 (0.02)
Male Ph vs Female Ph: $\frac{(T3+T4)-(T1+T2)}{2}$	-0.02 (0.04)	0.03 (0.02)	0.03 (0.02)	-0.02 (0.04)	0.03 (0.02)	0.03 (0.02)
<i>Pairwise Coeff Comparisons</i>						
T1 = T3 p	0.67	0.24	0.39	0.53	0.35	0.47
T2 = T4 p	0.69	0.31	0.17	0.96	0.20	0.09
T1 = T2 p	0.12	0.10	0.50	0.21	0.19	0.87
T3 = T4 p	0.11	0.12	0.22	0.06	0.07	0.20
Mean DV in PC	0.53	0.11	0.09	0.53	0.11	0.09
N	1013	1013	1013	1013	1013	1013

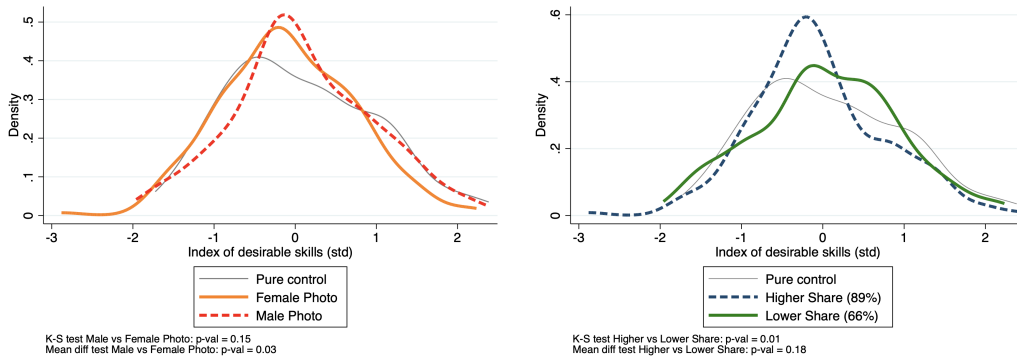
Robust standard errors in parentheses

Randomization inference p-values in square brackets

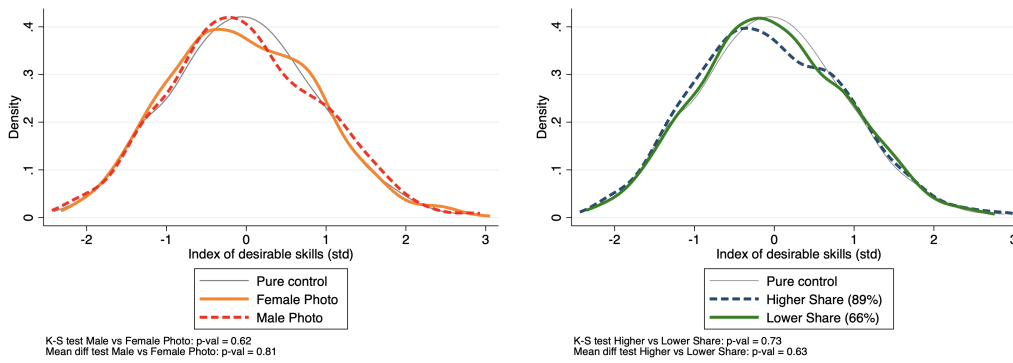
Note. The table shows OLS estimates for men only. The outcome variables are the same as in Table 2. The omitted category is the pure control group. Each regressor is a treatment dummy for the combination of a male (Male Ph) or female (Female Ph) photograph and information of a higher (Higher Sh) or lower (Lower Sh) past success. Regressions in Columns (1) to (3) control for covariates chosen using PDS Lasso, while Columns (4) to (6) show Double Machine Learning (DML) estimates. The panel “Overall effects” shows linear combinations of the coefficients and their standard errors to recover estimates of interest, e.g. the average effect of the Lower Share with respect to the pure control (“Lower Sh vs PC”) or the main effects shown in Table 2 (“Lower Sh vs Higher Sh”). The rows under “Pairwise Coeff Comparisons” show the p-values of pairwise tests of equality for the different coefficients in the table.

Figure A.4. Applicants' qualifications

(a) Men



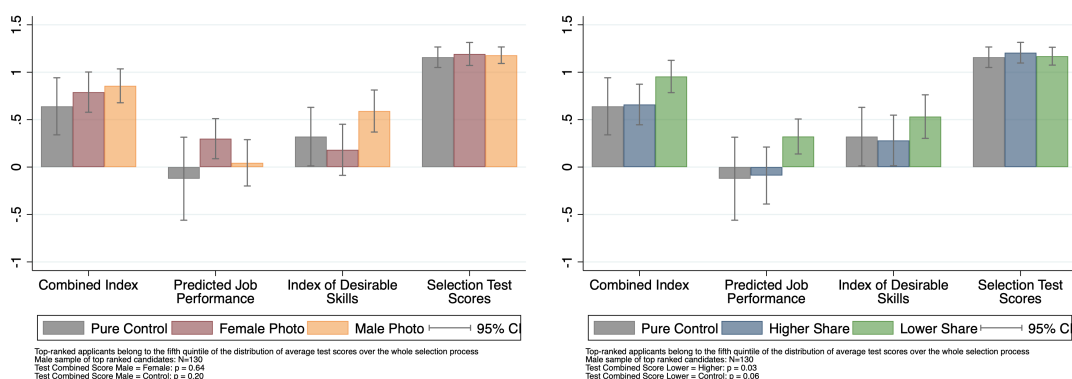
(b) Women



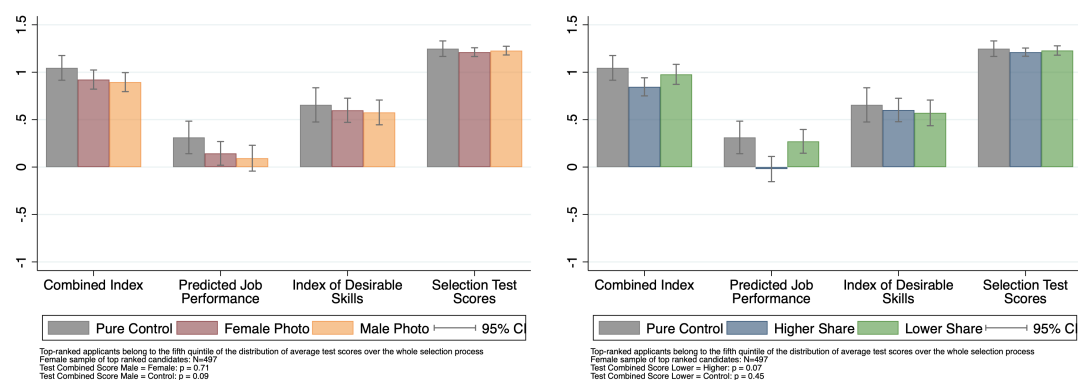
Note. The figure shows the distribution of male (Panel (a)) and female (Panel (b)) applicants' index of "desirable skills," residualized on stratification variables. The index is standardized and includes: receiving a first grade, being from a top UK university, volunteering frequently, cognitive skills, score in English pre-university tests, being in full time employment, having already graduated and studying a common subject for social work. The figure on the LHS shows the distributions by photograph treatment and the dashed red line is for the "Male Photo." The figure on the RHS shows the distributions by information treatment and the solid green line is for the "Lower Share." The solid grey line is for the pure control group. See Appendix B for details on the index construction.

Figure A.5. Quality among top-ranked applicants

(a) Men

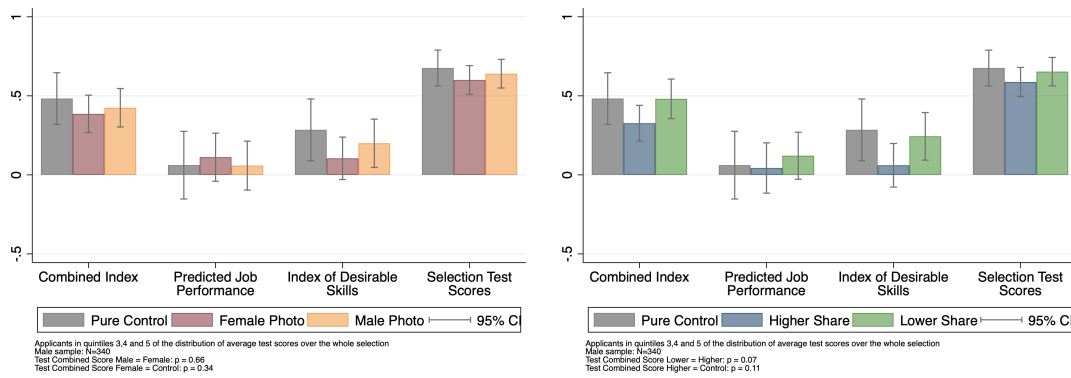


(b) Women



Note. The figure shows quality among top-ranked applicants, by photograph (LHS) and information treatment (RHS). Top-ranked applicants belong to the fifth quintile of average selection test-scores across the whole hiring process, pooling genders. There are three main proxies for quality in the Figure. “Predicted job performance” is calculated using a truncated linear regression in the pure control group on observable characteristics available at baseline (e.g., age, university ranking). The “Index of desirable skills” is defined as in Figure A.4. “Selection test scores” is the average obtained by the candidate across all the tests given by the employer in the hiring process. The combined index averages the three measures (all standardized to be mean 0 and unitary standard deviation), and re-standardizing the resulting average. See Appendix B for details on the indexes construction.

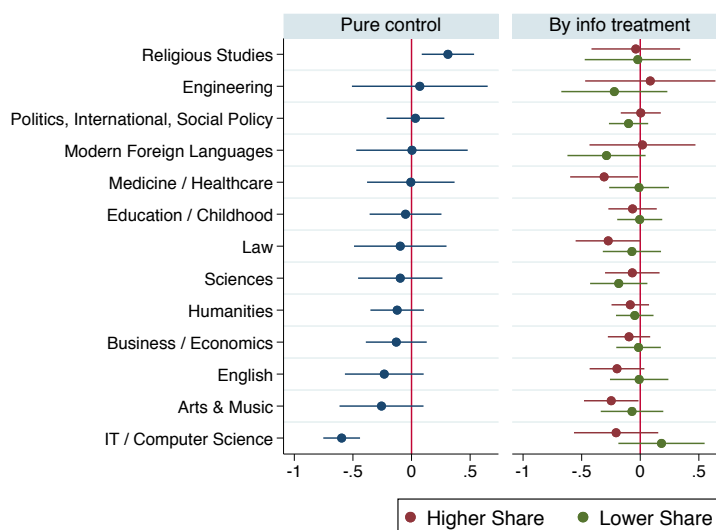
Figure A.6. Quality among top and mid-ranked applicants (men only)



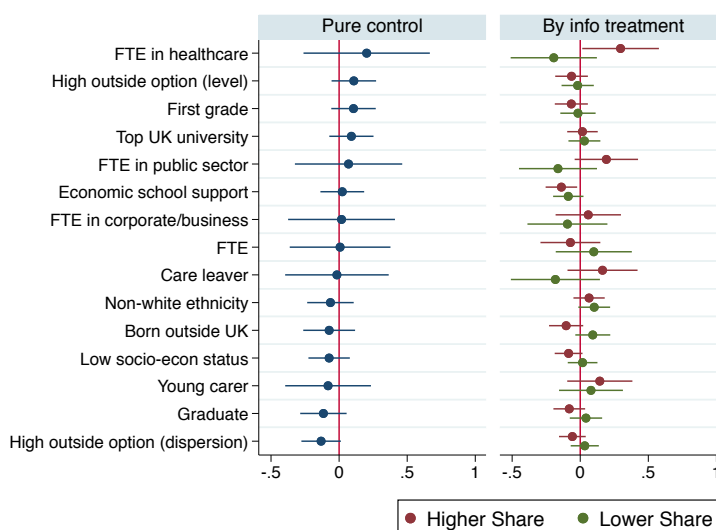
Note. The figure shows quality among top- and mid-ranked male applicants, by photograph treatment (LHS) and information treatment (RHS). Top- and mid-ranked applicants are defined as those belonging to the third, fourth and fifth quintile of the distribution of average test-scores across the whole hiring process and for both genders. These are the quintiles where all the offers are made (with 70% in the fifth one). There are three main proxies for quality in the figure, described in the notes for Figure A.4.

Figure A.7. Who are the men that apply?

(a) University subjects



(b) Background and experience



Note. The figure shows educational and demographic correlates of application behavior for men. Both panels (a) and (b) show the coefficients from three different OLS regressions - one for each experimental group (pure control, “Higher Share” and “Lower Share”) - of men’s application on a set of baseline variables. In Panel (a), the omitted category is “Psychology, Social Policy, Anthropology and Criminology,” which represents 33% of the sample. Panel (b) shows the relationship between other educational and background variables on the likelihood of applying. “High outside option (level)” and “High outside option (dispersion)” are indicators for having a predicted hourly wage and wage dispersion above the gender-specific median. Appendix B details the construction of both these measures. The other variables are described in the notes for Table 1.

Table A.7. On-the-job attitudes: additional variables

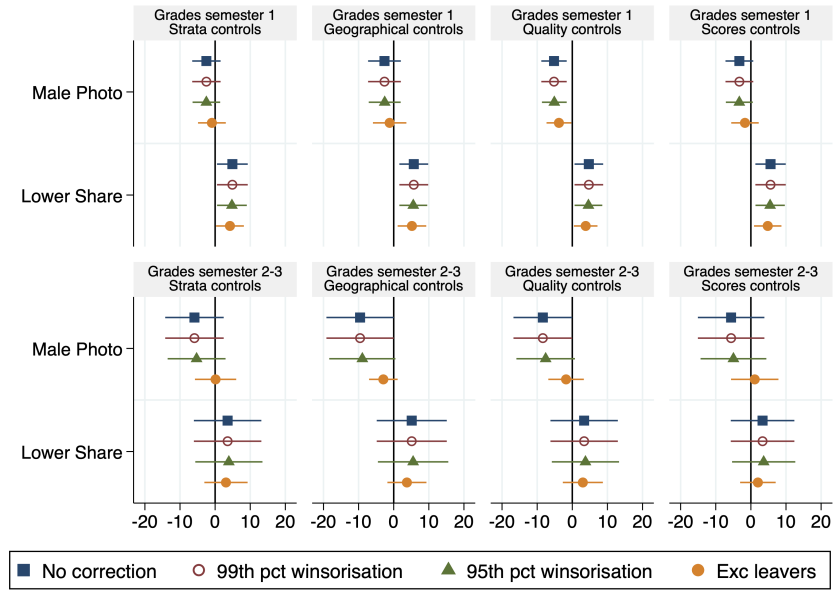
	(1)	(2)	(3)	(4)
	On-the-job attitudes			
	Concerned	Actual Workload	Job satisfaction	Perceived impact outside work
Panel A: Men				
Male Photo	0.115 (0.170) [0.45]	0.132 (0.203) [0.56]	0.319 (0.127) [0.04]	0.111 (0.124) [0.39]
Lower Share	-0.038 (0.164) [0.82]	0.023 (0.194) [0.90]	0.232 (0.094) [0.02]	0.121 (0.119) [0.31]
Photo=Info p	0.51	0.61	0.59	0.96
Mean DV	0.43	0.40	0.38	0.09
Mean DV in PC	0.47	0.55	0.59	0.25
N	43	46	112	66
Unique N	43	28	38	38
Panel B: Women				
Male Photo	-0.023 (0.072) [0.75]	0.059 (0.072) [0.40]	-0.011 (0.056) [0.84]	-0.012 (0.050) [0.80]
Lower Share	0.034 (0.072) [0.63]	0.049 (0.073) [0.54]	0.054 (0.055) [0.31]	-0.079 (0.051) [0.14]
Photo=Info p	0.57	0.92	0.41	0.41
Mean DV	0.56	0.33	0.61	0.17
Mean DV in PC	0.50	0.32	0.68	0.29
N	191	198	487	295
Unique N	191	126	175	174
<i>Gender Differences</i>				
Photo: p Men=Women	0.43	0.72	0.01	0.34
Info: p Men=Women	0.67	0.89	0.09	0.11

Robust standard errors in parentheses
 $p < 0.05$, $p < 0.01$, $p < 0.001$

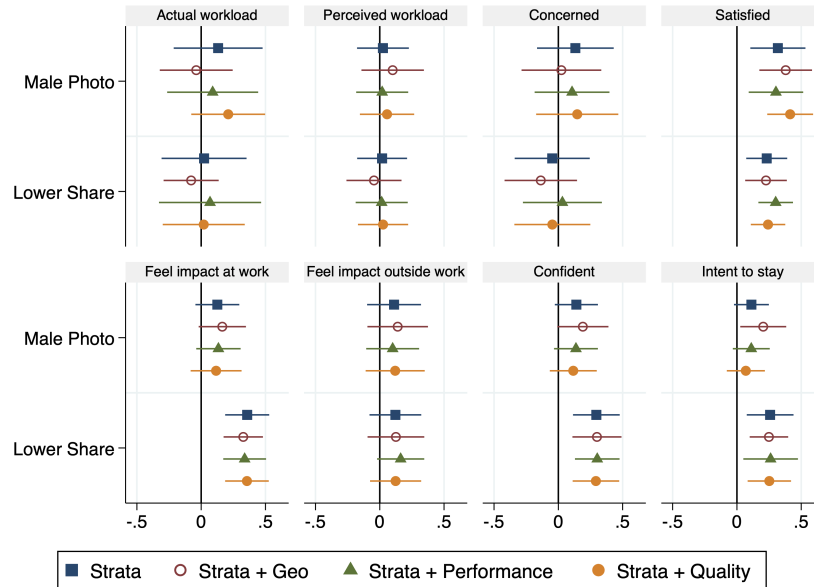
Note. The table shows OLS estimates with controls for stratification variables and survey wave. “Concerned” is an indicator variable equal to one if a worker feels any personal, financial, work, well-being or health-related concern right before starting the job. “Actual Workload” is an indicator variable equal to one if a worker has a workload of 17 or more cases. “Job Satisfaction” is a dummy for whether the person feels satisfied with their work. “Perceived impact outside work” is an indicator equal to one if a worker feels that s/he is having positive social impact outside work. Data for Columns (2) to (4) are from the organization’s surveys, conducted every semester. The question on concerns in Column (1) was asked only once when a candidate accepted the job offer. For an additional description of this table, see the notes for Table 3.

Figure A.8. Men’s on-the-job outcomes: robustness to different controls and outliers

(a) On-the-job performance



(b) Survey attitudes



Note. The figure shows robustness exercises for the coefficients presented in Tables 3 and A.7: panel (a) for grade outcomes (Columns (1) and (2) of Table 3 and Table A.7). In each subgraph of panel (a), the four coefficients on a treatment dummy correspond to four regressions, which differ in whether the outcome is not corrected (square), winsorized at the 99th percentile (hollow circle), at the 95th percentile (triangle) or excludes job leavers (circle). The subgraphs in each row differ in the controls used: i) strata only, ii) strata and geographical controls (region fixed effects, whether a person was assigned to the preferred region and whether they applied previously), iii) strata and the index of “desirable skills” defined in Figure A.4 and iv) strata and the test score achieved in the last stage of the hiring process. Each subgraph of panel (b) reports results from four different regressions which differ in the controls used, but have the same outcome variable (reported at the top of the subgraph). The four set of controls are the same as in panel (a). Error bars show 90% confidence intervals in both panels (a) and (b).

Table A.8. Application outcomes for women: fully interacted treatments

	(1)	(2)	(3)	(4)	(5)	(6)
	PDS Lasso Controls			DML Controls		
	Applied	Offer	Accepted	Applied	Offer	Accepted
T1: Female Ph + Higher Sh	0.011 (0.023) [0.64]	-0.005 (0.013) [0.71]	-0.016 (0.011) [0.13]	0.015 (0.023) [0.51]	-0.004 (0.013) [0.65]	-0.015 (0.011) [0.17]
T2: Female Ph + Lower Sh	0.034 (0.023) [0.11]	-0.006 (0.013) [0.66]	-0.014 (0.011) [0.22]	0.036 (0.023) [0.14]	-0.004 (0.013) [0.74]	-0.010 (0.011) [0.31]
T3: Male Ph + Higher Sh	-0.001 (0.023) [0.97]	-0.005 (0.013) [0.72]	-0.001 (0.011) [0.94]	-0.001 (0.023) [0.99]	-0.006 (0.013) [0.45]	-0.001 (0.011) [0.52]
T4: Male Ph + Lower Sh	-0.055 (0.023) [0.02]	-0.007 (0.013) [0.60]	-0.005 (0.011) [0.66]	-0.056 (0.023) [0.02]	-0.004 (0.013) [0.69]	-0.004 (0.011) [0.75]
<i>Overall effects</i>						
Lower Sh vs PC: $\frac{T2+T4}{2}$	-0.01 (0.02)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.02)	-0.00 (0.01)	-0.01 (0.01)
Lower Sh vs Higher Sh: $\frac{(T2+T4)-(T1+T3)}{2}$	-0.02 (0.02)	-0.00 (0.01)	-0.00 (0.01)	-0.02 (0.02)	0.00 (0.01)	0.00 (0.01)
Male Ph vs PC: $\frac{T3+T4}{2}$	-0.03 (0.02)	-0.01 (0.01)	-0.00 (0.01)	-0.03 (0.02)	-0.01 (0.01)	-0.00 (0.01)
Male Ph vs Female Ph: $\frac{(T3+T4)-(T1+T2)}{2}$	-0.05 (0.02)	0.00 (0.01)	0.01 (0.01)	-0.05 (0.02)	-0.00 (0.01)	0.01 (0.01)
<i>Pairwise Coeff Comparisons</i>						
T1 = T3 p	0.61	0.98	0.15	0.49	0.89	0.21
T2 = T4 p	0.00	0.97	0.43	0.00	0.99	0.61
T1 = T2 p	0.32	0.90	0.80	0.36	0.96	0.63
T3 = T4 p	0.02	0.89	0.70	0.02	0.85	0.79
Mean DV in PC	0.59	0.09	0.06	0.59	0.09	0.06
N	4404	4404	4404	4404	4404	4404

Robust standard errors in parentheses
Randomization inference p-values in square brackets

Note. The table shows OLS estimates for women only. See the notes for Table A.6 for a detailed description of this table.

Table A.9. Gender differences in the reaction to treatments

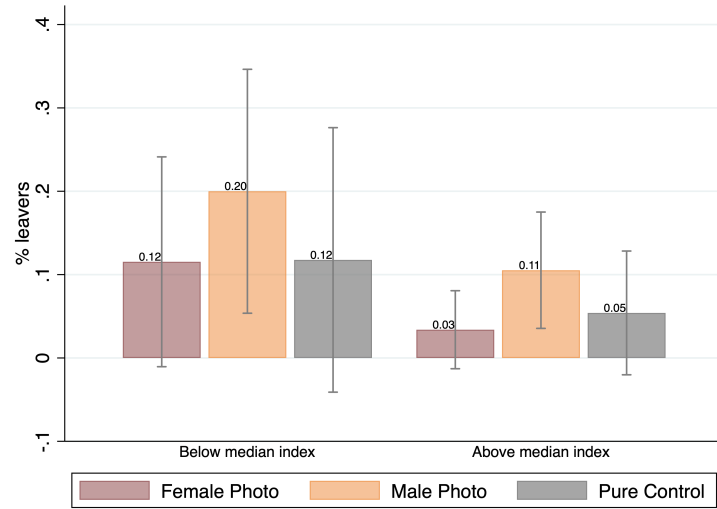
	(1)	(2)	(3)	(4)	(5)
DV: Applied & never drop-out					
Male x Male Photo	0.035 (0.038)	0.043 (0.039)	0.046 (0.039)	0.030 (0.040)	0.032 (0.041)
Male x Lower Share	0.086 (0.038)	0.090 (0.038)	0.096 (0.038)	0.083 (0.039)	0.098 (0.040)
N	4320	4320	4320	4320	4320
<i>Controls</i>					
Full Controls*Gender		✓	✓	✓	✓
Subject FE*Gender			✓		✓
Full Controls*Treats				✓	✓
Subject FE*Treats					✓

Standard errors in parentheses

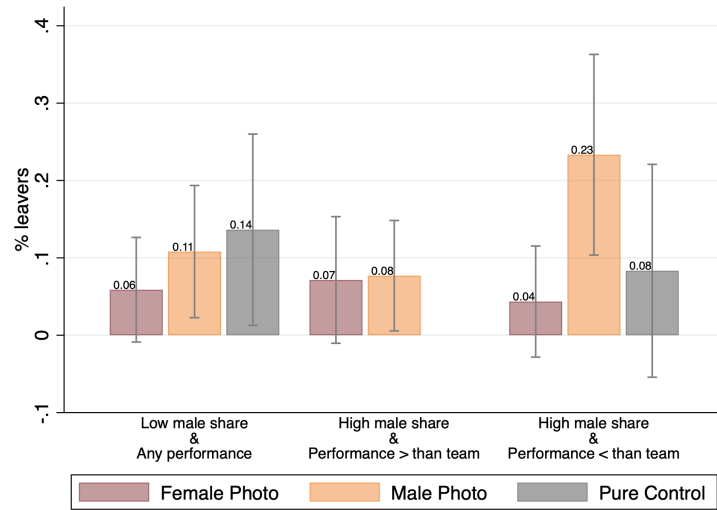
Note. The table shows regressions in which the dummy for applying is regressed on the treatment variables “Male Photo” and “Lower Share,” on an indicator for being a male candidate and on the interactions between each treatment and the male indicator. The set “Full Controls” includes all the variables presented in Table 1. “Subject FE” are a set of 15 indicators for the subject studied in university. “Full Controls*Gender” indicates that the controls are added to the regression and are also interacted with the male dummy. “Full Controls*Treats” indicates that the controls are added to the regression and are also interacted with each of the two treatment dummies. Column (1) only controls for stratification variables.

Figure A.9. Female turnover

(a) By index of desirable skills



(b) By team male share and relative performance



Note. Panel (a) shows female turnover by observable skills above or below median of the index defined in Figure A.4. Panel (b) splits the sample in three categories: i) women allocated in teams with median or below median male share ($\leq 20\%$) and any relative performance, ii) women allocated in teams with higher than median male share ($> 20\%$) and with individual performance which is better than the leave-one-out team average in the first semester and iii) women allocated in teams with higher than median male share ($> 20\%$) and with individual performance which is equal or worse than the leave-one-out team average in the first semester.

Table A.10. Overall gender shares and performance

	Male Share			Share of Top Performers		
	Applicants	Offerees	Workers	Women	Men	Overall
	<i>Photographs</i>					
Female Photo	17%	15%	16%	52%	63%	54%
Male Photo	18%	21%	21%	59%	42%	55%
	<i>Information</i>					
Higher Share	16%	14%	14%	54%	40%	52%
Lower Share	19%	22%	23%	58%	56%	58%
Pure Control	17%	23%	21%	54%	54%	54%

Note. The first three columns of this table show the male share among applicants (Column 1), people who received a job offer (Column 2) and workers (Column 3), excluding people who drop out before completing the program. The last three columns show the share of top performers (people with an average score above 60%) by women (Column 4), men (Column 5) and for the pooled sample, excluding people who do not complete the program.

Figure A.10. Survey questions about the trade-off between retention and performance

(a) Vignette question 1

Scenario A	Scenario B
<ul style="list-style-type: none"> ■ Starting cohort of 100 people ■ Everyone completes the programme (2 full years) ■ 51% are top performers by the end of the programme 	<ul style="list-style-type: none"> ■ Starting cohort of 100 people ■ 18 people drop-out before finishing the programme ■ 61% are top performers by the end of the programme

(b) Vignette question 2

Scenario A	Scenario B
<ul style="list-style-type: none"> ■ Starting cohort of 100 people ■ Everyone completes the programme (2 full years) ■ 51% are top performers by the end of the programme ■ Women are 85% by the end of the programme 	<ul style="list-style-type: none"> ■ Starting cohort of 100 people ■ 18 people drop-out before finishing the programme ■ 61% are top performers by the end of the programme ■ Women are 71% by the end of the programme

Note. The figure shows the two main questions asked in the survey with the partner organization about the trade-off between retention and performance. The text of both questions was the following: *Consider the following two scenarios. Consider them identical besides the characteristics given below. Which of these scenarios would you prefer for the local communities where the organization operates? In both, consider as top performer somebody with commendable or excellent average scores ($\geq 60\%$) in the practice tests of the program.* The implementation of the survey is described in Appendix C.

Table A.11. Correlation between exposure to gender occupational segregation and confidence in beliefs on social workers' performance

DV: Confidence in beliefs of performance in social work		
	(1)	(2)
Sample:	Men	Women
Exposure to high gender segregation	-7.294 (2.915)	-1.895 (2.121)
N	146	340
Mean DV	74.66	80.18

Robust standard errors in parentheses

Note. The dependent variable is constructed from the questions: "On a scale from 0 to 100, how confident are you of your answer [about the performance of a man/woman in social work]?". "Exposure to high gender segregation" is equal to one if the Duncan index of occupational gender segregation in the postcode where a respondent was living when she/he was 14 years old is above the median of the sample. The regression controls for ethnicity, survey sample and the respondent's beliefs on the performance of a man or woman in social work. Data are from the out-of-trial surveys described in Section V. The sample includes people whose postcode could be matched with the 2011 Census.

Table A.12. Heterogeneous treatment effects: controls for strata only

	(1)	(2)	(3)	(4)	(5)	(6)
	DV: Applied & no drop-out					
	<i>By background</i>				<i>By outside option</i>	
	Job Genderization		Familiarity with Job		Wage Dispersion	
	Low	High	Low	High	Low	High
Panel A: Men						
Male Photo	-0.029 (0.049) [0.56]	-0.006 (0.051) [0.89]	-0.050 (0.057) [0.41]	0.002 (0.044) [0.96]	-0.015 (0.045) [0.73]	-0.023 (0.055) [0.67]
Lower Share	0.011 (0.049) [0.83]	0.129 (0.051) [0.01]	0.170 (0.057) [0.00]	0.011 (0.044) [0.82]	0.030 (0.045) [0.54]	0.130 (0.055) [0.02]
Photo=Info p	0.56	0.07	0.01	0.89	0.49	0.05
Mean DV	0.59	0.45	0.44	0.57	0.56	0.45
Mean DV in PC	0.57	0.51	0.46	0.57	0.58	0.46
N	400	379	297	510	483	324
Photo: High=Low p		0.74		0.46		0.91
Info: High=Low p		0.09		0.03		0.16
Panel B: Women						
Male Photo	-0.069 (0.024) [0.00]	-0.040 (0.024) [0.09]	-0.013 (0.028) [0.68]	-0.074 (0.020) [0.00]	-0.058 (0.019) [0.01]	-0.031 (0.034) [0.37]
Lower Share	-0.017 (0.024) [0.48]	-0.016 (0.024) [0.48]	-0.048 (0.028) [0.08]	-0.001 (0.020) [0.98]	-0.020 (0.019) [0.33]	-0.002 (0.034) [0.94]
Photo=Info p	0.12	0.47	0.39	0.01	0.16	0.53
Mean DV	0.62	0.60	0.55	0.63	0.61	0.57
Mean DV in PC	0.57	0.61	0.56	0.61	0.60	0.55
N	1710	1693	1240	2273	2644	869
Photo: High=Low p		0.39		0.08		0.48
Info: High=Low p		0.98		0.18		0.65
<i>Gender differences</i>						
Photo: Men=Women p	0.47	0.55	0.55	0.12	0.38	0.91
Info: Men=Women p	0.61	0.01	0.00	0.82	0.31	0.04

Note. The table replicates the results of Table 4, but with only controls for stratification variables (ethnicity and early registration). See the notes for Table 2 for a detailed description of this table.

Table A.13. Additional heterogeneous treatment effects

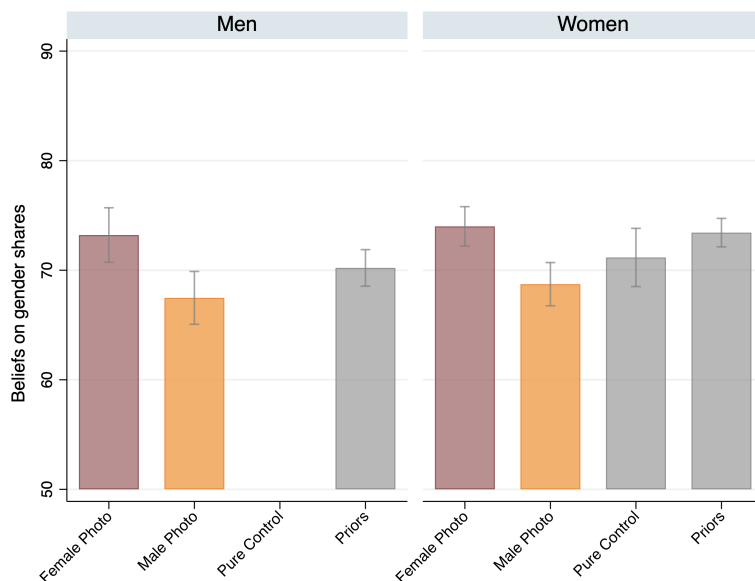
	(1)	(2)	(3)	(4)
	Applied & no drop-out			
	Outside Option Wage		Predicted On-the-job Score	
	Low	High	Low	High
Panel A: Men				
Male Photo	-0.031 (0.048) [0.53]	0.019 (0.047) [0.70]	-0.006 (0.043) [0.88]	-0.017 (0.062) [0.80]
Lower Share	0.043 (0.049) [0.38]	0.099 (0.049) [0.04]	0.041 (0.043) [0.35]	0.118 (0.061) [0.06]
Photo=Info p	0.28	0.24	0.45	0.12
Mean DV	0.56	0.47	0.53	0.49
Mean DV in PC	0.49	0.57	0.52	0.54
N	395	412	540	267
Photo: High=Low p		0.47		0.88
Info: High=Low p		0.42		0.30
Panel B: Women				
Male Photo	-0.032 (0.023) [0.15]	-0.072 (0.024) [0.00]	-0.060 (0.020) [0.00]	-0.036 (0.031) [0.21]
Lower Share	-0.029 (0.023) [0.24]	-0.003 (0.024) [0.90]	-0.050 (0.020) [0.02]	0.054 (0.029) [0.06]
Photo=Info p	0.91	0.05	0.73	0.03
Mean DV	0.64	0.56	0.63	0.54
Mean DV in PC	0.62	0.55	0.57	0.63
N	1769	1744	2340	1173
Photo: High=Low p		0.23		0.48
Info: High=Low p		0.44		0.00
<i>Gender differences</i>				
Photo: Men=Women p	0.99	0.09	0.25	0.78
Info: Men=Women p	0.19	0.06	0.05	0.34

Robust standard errors in parentheses
Randomization Inference p-values in square brackets

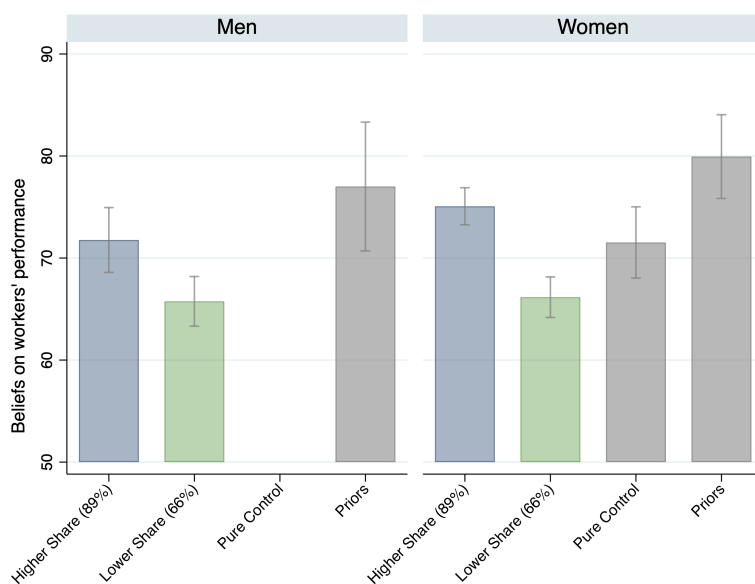
Note. The table shows OLS estimates with controls for stratification variables. The outcome variable is the indicator for applying and never withdrawing from the process. Columns (1) and (2) split the sample by below- or above-median predicted hourly wage in the UK labor market. Columns (3) and (4) split the sample below or above the 75th percentile of predicted performance on the job. The construction of both these variables is described in Appendix B. For an additional description of this table, see the notes for Table 4.

Figure A.11. Out-of-trial surveys: comparison of beliefs across treatment groups versus priors

(a) Gender Shares



(b) Share of high-performing workers



Note. The figure uses data from the out-of-trial surveys. Assignment to the "Pure control" was implemented only for women from the applicants' sample ($N=61$). The first three bars of Panel (a) show beliefs on the share of female applicants for the job offered by the partner organization in the "Female Photo," "Male Photo" and pure control, respectively. The last bar shows the average answer to the question: "Out of 100 workers in social work, what is the number of female workers in your opinion?", which was asked before treatment assignment. In Panel (b), the first three bars show beliefs on the share of workers who are successful on-the-job in the "Higher Share," "Lower Share" and pure control, respectively. The last bar shows the share of respondents who think that on average workers achieve a performance of more than 65% on the job. This is constructed from the questions "On a scale from 0 to 100, what do you think is the performance of a [man/woman] in social work?", which were asked before treatment assignment. I compute the weighted average of answers (with weights given by the actual gender shares in social work in the UK) and code as one averages above 65.

Table A.14. Out-of-trial surveys: summary statistics, differences between treatment groups and MHT corrections by gender

Used for interpretation of:	Variable	Mean	SD	Female - Male Photo			Higher (89%) - Lower (66%) Info				
				Diff	P-values		Diff	P-values			
					Unadjusted	FDR		FWER	Unadjusted	FDR	FWER
Panel A: Men											
Photographs	% female applicants	70.38	11.96	5.73	0.00	0.01	0.05	0.65	0.72	1.00	1.00
	Employer wants your gender	57.30	49.60	-43.99	0.00	0.00	0.00	-0.33	0.96	1.00	0.96
	Discrimination by customers	47.75	50.09	18.03	0.02	0.09	0.29	-7.77	0.30	1.00	1.00
	Job desirable for men	61.24	48.86	-6.99	0.34	0.68	1.00	-8.37	0.25	1.00	1.00
	Job desirable for women	93.26	25.14	6.89	0.07	0.25	0.85	-4.19	0.27	1.00	1.00
Information	% of high-perf applicants	70.73	19.67	-4.41	0.13	0.41	0.97	8.55	0.00	0.03	0.06
	% of high-performing workers	68.83	13.82	0.70	0.74	0.99	1.00	6.01	0.00	0.03	0.09
	Self-reported ability	6.56	2.36	-0.11	0.75	0.99	1.00	-0.04	0.92	1.00	1.00
	Own exp ranking for the job	53.57	25.74	4.73	0.22	0.56	0.99	6.54	0.09	0.56	0.91
	Employer's exp ranking for the job	49.62	25.80	2.51	0.52	0.90	1.00	0.37	0.93	1.00	1.00
	Job difficulty	61.28	21.26	4.34	0.17	0.48	0.98	1.37	0.67	1.00	1.00
	Quality standards	72.76	21.49	8.86	0.06	0.25	0.81	3.21	0.50	1.00	1.00
	Promotion difficulty	54.49	19.01	-3.29	0.25	0.56	1.00	0.16	0.96	1.00	1.00
	Application effort	60.55	21.45	5.21	0.28	0.58	1.00	-2.75	0.57	1.00	1.00
	Number of applicants	48.99	22.75	0.82	0.81	0.99	1.00	3.56	0.30	1.00	1.00
Other explanations	Wage level	43.13	18.88	-0.45	0.88	0.99	1.00	5.59	0.07	0.56	0.86
	Job high social status	47.19	50.06	-1.06	0.89	0.99	1.00	6.87	0.36	1.00	1.00
	Felt fooled by ad	28.65	45.34	0.48	0.94	0.99	1.00	2.08	0.76	1.00	1.00
Panel B: Women											
Photographs	% female applicants	71.32	12.48	5.27	0.00	0.00	0.00	1.82	0.19	0.37	0.99
	Employer wants your gender	71.69	45.12	29.91	0.00	0.00	0.00	1.06	0.83	0.75	1.00
	Discrimination by customers	42.59	49.52	4.76	0.39	0.97	1.00	11.64	0.03	0.11	0.65
	Job desirable for men	72.84	44.55	-17.96	0.00	0.00	0.00	0.90	0.86	0.75	1.00
	Job desirable for women	87.04	33.64	13.26	0.00	0.00	0.01	6.01	0.11	0.23	0.94
Information	% of high-perf applicants	75.35	18.81	2.37	0.26	0.87	1.00	9.65	0.00	0.00	0.00
	% of high-performing workers	70.61	13.06	1.07	0.46	0.97	1.00	8.91	0.00	0.00	0.00
	Self-reported ability	7.76	1.80	-0.22	0.27	0.87	1.00	-0.49	0.01	0.06	0.36
	Own exp ranking for the job	49.23	20.43	2.33	0.30	0.90	1.00	-2.88	0.20	0.37	1.00
	Employer's exp ranking for the job	50.95	21.48	-0.10	0.97	1.00	0.97	-0.76	0.75	0.73	1.00
	Job difficulty	64.83	18.73	-0.49	0.82	1.00	1.00	-7.27	0.00	0.00	0.01
	Quality standards	76.74	19.75	-4.98	0.16	0.62	0.99	-2.27	0.53	0.69	1.00
	Promotion difficulty	56.02	15.76	0.27	0.88	1.00	1.00	1.89	0.28	0.45	1.00
	Application effort	59.58	23.40	-3.92	0.34	0.92	1.00	2.78	0.50	0.69	1.00
	Number of applicants	57.76	19.60	0.51	0.82	1.00	1.00	-1.33	0.54	0.69	1.00
Other explanations	Wage level	49.01	16.70	-0.28	0.91	1.00	1.00	-1.25	0.60	0.69	1.00
	Job high social status	56.48	49.65	2.01	0.72	1.00	1.00	11.19	0.04	0.11	0.71
	Felt fooled by ad	29.54	45.69	7.06	0.16	0.62	0.99	2.64	0.60	0.69	1.00

Note. The table shows summary statistics and differences between treatment groups in the main outcomes measured in the out-of-trial surveys. Columns denoted with “FDR” and “FWER” report multiplicity-adjusted p-values following [Benjamini et al. \(2006\)](#) and [List et al. \(2019\)](#), within gender and treatment. Variables are divided in three categories depending on whether they are discussed in relation to the interpretation of the photographs (section V.A), information (section V.B) or alternative interpretations (section F) in the main body of the paper. A detailed description of the variables and corresponding survey questions can be found in Appendix C.

Table A.15. Out-of-trial surveys: summary statistics, differences between treatment groups and MHT corrections by sample

Used for interpretation of:	Variable	Mean	SD	Female - Male Photo			Higher (89%) - Lower (66%) Info				
				Diff	P-values		Diff	P-values		FDR	FWER
					FDR	FWER		FDR	FWER		
Panel A: 2018 Applicants											
Photographs	% female applicants	69.84	13.16	5.28	0.00	0.01	0.03	1.32	0.44	0.66	1.00
	Employer wants own gender	40.91	49.27	-28.25	0	0	0	1.8	0.78	1.00	1.00
	Discrimination by customers	46.06	49.95	9.51	0.14	0.27	0.97	13.73	0.03	0.11	0.63
	Job desirable for men	72.20	44.89	-19.32	0.00	0.00	0.00	2.72	0.64	0.84	1.00
	Job desirable for women	84.23	36.52	23.11	0.00	0.00	0.00	6.77	0.15	0.35	0.99
Information	% of high-perf applicants	76.48	20.17	1.73	0.51	0.62	1.00	7.64	0.00	0.03	0.09
	% of high-performing workers	70.98	13.33	3.21	0.06	0.15	0.82	5.52	0.00	0.02	0.04
	Self-reported ability	8.14	1.36	-0.12	0.49	0.62	1.00	-0.28	0.12	0.34	0.96
	Own exp ranking for the job	53.28	21.24	6.63	0.01	0.04	0.33	-1.79	0.51	0.75	1.00
	Employer's exp ranking for the job	52.38	21.79	1.59	0.57	0.62	1.00	-1.04	0.71	0.90	1.00
	Job difficulty	63.05	19.71	0.08	0.97	0.95	1.00	-5.49	0.03	0.11	0.60
	Quality standards	75.60	20.56	-4.14	0.38	0.62	1.00	-4.46	0.35	0.56	1.00
	Promotion difficulty	55.72	16.91	1.59	0.47	0.62	1.00	0.52	0.81	1.00	1.00
	Application effort	60.96	24.49	-1.62	0.77	0.73	1.00	2.77	0.62	0.84	1.00
	Number of applicants	60.02	18.56	1.16	0.63	0.65	1.00	-3.36	0.16	0.35	0.99
Other explanations	Wage level	51.23	14.27	1.82	0.56	0.62	1.00	0.18	0.95	1.00	0.96
	Job high social status	59.34	49.22	7.96	0.21	0.39	0.99	16.94	0.01	0.04	0.24
	Felt fooled by ad	33.47	47.29	9.64	0.11	0.24	0.95	6.89	0.26	0.53	1.00
Panel B: Prolific sample											
Photographs	% female applicants	72.10	11.37	5.62	0.00	0.00	0.00	1.56	0.27	1.00	1.00
	Employer wants own gender	36.4	48.21	-40.33	0	0	0	-2.58	0.67	1.00	1.00
	Discrimination by customers	42.75	49.57	9.16	0.14	0.76	0.96	-3.71	0.55	1.00	1.00
	Job desirable for men	65.27	47.70	-9.92	0.09	0.75	0.92	-7.87	0.18	1.00	0.99
	Job desirable for women	93.89	23.99	0.00	1.00	1.00	1.00	-1.43	0.63	1.00	1.00
Information	% of high-perf applicants	71.26	18.03	-1.52	0.50	1.00	1.00	10.81	0.00	0.00	0.00
	% of high-performing workers	69.06	13.30	-1.16	0.48	1.00	1.00	10.01	0.00	0.00	0.00
	Self-reported ability	6.59	2.36	-0.25	0.39	1.00	1.00	-0.40	0.17	1.00	0.99
	Own exp ranking for the job	48.44	23.41	0.13	0.97	1.00	1.00	2.62	0.37	1.00	1.00
	Employer's exp ranking for the job	48.77	24.11	0.22	0.94	1.00	1.00	0.34	0.91	1.00	1.00
	Job difficulty	64.05	19.71	2.17	0.37	1.00	1.00	-3.10	0.20	1.00	1.00
	Quality standards	75.00	20.49	3.06	0.40	1.00	1.00	2.38	0.51	1.00	1.00
	Promotion difficulty	55.26	17.05	-3.40	0.11	0.75	0.95	1.91	0.37	1.00	1.00
	Application effort	59.36	21.52	0.00	1.00	1.00	1.00	-0.69	0.85	1.00	1.00
	Number of applicants	49.73	22.20	0.06	0.98	1.00	1.00	3.73	0.17	1.00	0.99
Other explanations	Wage level	44.95	18.63	-1.34	0.56	1.00	1.00	2.00	0.39	1.00	1.00
	Job high social status	47.71	50.04	-5.34	0.39	1.00	1.00	3.09	0.62	1.00	1.00
	Felt fooled by ad	25.19	43.49	0.00	1.00	1.00	1.00	-1.91	0.72	1.00	1.00

Note. The table shows summary statistics and differences between treatment groups in the main outcomes measured in the out-of-trial surveys, by survey sample. Columns denoted with “FDR” and “FWER” report multiplicity-adjusted p-values following [Benjamini et al. \(2006\)](#) and [List et al. \(2019\)](#), within gender and treatment. Variables are divided in three categories depending on whether they are discussed in relation to the interpretation of the photographs (section [V.A](#)), information (section [V.B](#)) or alternative interpretations (section [F](#)) in the main body of the paper. A detailed description of the variables and corresponding survey questions can be found in [Appendix C](#).

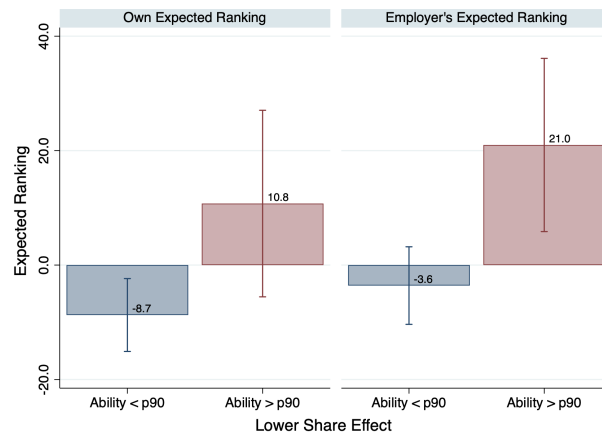
Table A.16. Out-of-trial surveys: updating of expected ranking by ability

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	DV: Own Expected ranking								
Ability proxy:	Predicted Job Score				Self-report				Self-rep + Pred
	50th p	50th p g spec	75th p	90th p	mean	mean g spec	50th p	50th p g spec	50th p
Panel A: Men									
Low Ability: Δ 66 - 89 %	-10.20 (5.20)	-10.68 (5.17)	-8.24 (4.22)	-8.73 (3.87)	-9.23 (4.80)	-8.38 (6.97)	-6.64 (3.95)	-9.23 (4.98)	-6.33 (3.77)
High Ability: Δ 66 - 89 %	-2.99 (5.14)	-2.32 (5.13)	-1.38 (7.00)	10.75 (9.91)	-3.59 (5.47)	-5.40 (4.08)	-7.01 (8.67)	-3.59 (5.47)	-4.70 (11.06)
DID	7.20	8.36	6.86	19.48	5.64	2.98	-0.37	5.64	1.63
DID p	0.33	0.25	0.40	0.07	0.44	0.70	0.97	0.44	0.89
Observations	178	178	178	178	178	178	178	178	178
Panel B: Women									
Low Ability: Δ 66 - 89 %	3.78 (3.08)	3.78 (3.08)	2.65 (2.67)	2.82 (2.34)	-2.36 (3.31)	-2.36 (3.31)	0.38 (2.56)	0.38 (2.56)	1.29 (2.42)
High Ability: Δ 66 - 89 %	1.74 (3.37)	1.74 (3.37)	3.16 (4.99)	0.72 (9.20)	5.55 (2.94)	5.55 (2.94)	7.40 (4.42)	7.40 (4.42)	10.54 (6.34)
DID	-2.04	-2.04	0.51	-2.11	7.90	7.90	7.02	7.02	9.25
DID p	0.66	0.66	0.93	0.83	0.07	0.07	0.16	0.16	0.18
Observations	324	324	324	324	325	325	325	325	324

Bootstrapped standard errors in parentheses

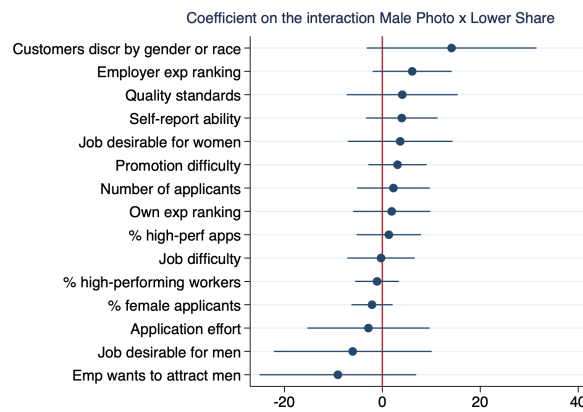
Note. The table shows robustness exercises for the RHS of Figure 5 through regressions of a respondent’s own expected ranking for the job on the interaction between treatment dummies and an indicator for “High Ability.” The table reports the difference in the coefficients on the “Lower Share” minus “Higher Share” treatment, separately by respondents of high and low ability. All regressions control for the survey sample. In Columns (1) to (4), “High Ability” indicates a respondent above the 50th, 75th or 90th percentile of the distribution of predicted job performance (computed using the method described in Appendix B). In Columns (5) to (8), “High Ability” indicates survey respondents with self-reported ability above mean or median, overall or gender-specific. Column (9) defines “High Ability” respondents above median in both self-reports and predicted job performance. “DID” reports the difference-in-difference between the two information treatments and high- versus low-ability respondents, and “DID p” the corresponding p-value. Data are from the out-of-trial surveys described in Section V.

Figure A.12. Out-of-trial surveys: updating of beliefs on expected ranking by ability



Note. The figure reports coefficients from a regression of expected ranking on the “Lower Share” dummy, an indicator for being above the 90th percentile of the distribution of predicted job performance and their interaction. The panel on the LHS uses the respondent’s own expected ranking as dependent variable, and the one on the RHS the employer’s expected ranking. The blue bars show the effect of the “Lower Share” for people below the 90th percentile and the red bars show the for people above the 90th percentile, both compared to people in the “Higher Share” with the same ability level.

Figure A.13. Out-of-trial surveys: interaction between photographs and information



Note. The figure shows the coefficient on the interaction between the “Male Photo” and the “Lower Share” indicators, for each of the given row variables and pooling genders. Data are from the out-of-trial surveys described in Section V.

Table A.17. Employer's hiring criteria

VARIABLES	(1)	(2)	(3)	(4)
	Information		Photographs	
	Offer	p-val	Offer	p-val
Top University * T^1	0.079 (0.025)		0.095 (0.024)	
Top University * T^2	0.119 (0.026)	0.27	0.100 (0.028)	0.9
First Grade * T^1	0.109 (0.031)		0.069 (0.030)	
First Grade * T^2	0.107 (0.032)	0.96	0.153 (0.033)	0.06
Common Subject * T^1	0.007 (0.019)		0.023 (0.019)	
Common Subject * T^2	0.047 (0.020)	0.06	0.034 (0.020)	0.61
Past Volunteering * T^1	0.033 (0.020)		0.050 (0.020)	
Past Volunteering * T^2	0.046 (0.020)	0.65	0.027 (0.021)	0.41
Math Pre-Uni Score * T^1	0.025 (0.028)		-0.001 (0.025)	
Math Pre-Uni Score * T^2	-0.015 (0.026)	0.29	0.014 (0.030)	0.69
English Pre-Uni Score * T^1	0.076 (0.024)		0.075 (0.024)	
English Pre-Uni Score * T^2	0.051 (0.025)	0.47	0.051 (0.026)	0.49
FTE * T^1	0.127 (0.024)		0.094 (0.023)	
FTE * T^2	0.074 (0.023)	0.11	0.104 (0.024)	0.75
Graduate * T^1	0.042 (0.025)		0.070 (0.025)	
Graduate * T^2	0.100 (0.026)	0.10	0.077 (0.025)	0.83
Observations	2,271		2,271	

Robust standard errors in parentheses

Note. The dependent variable is an indicator equal to one if a person got the job offer, conditional on applying. The independent variables interact the different treatments arms with observable characteristics of the applicants. In Column (1), T^2 indicates the lower share information (and T^1 the alternative information). In Column (3), T^2 indicates a male photograph (and T^1 a female photograph). All regressions control for stratification variables (gender, ethnicity and early registration). Columns (2) and (4) report the p-value of tests of equality of the coefficients of a certain observable characteristic interacted with T^1 versus T^2 . "Past Volunteering" is equal to one if the candidate volunteered frequently in the past. "Math Pre-Uni Score" and "English Pre-Uni Score" are equal to one if the candidate took the highest grade in Math and English pre-university qualifications.

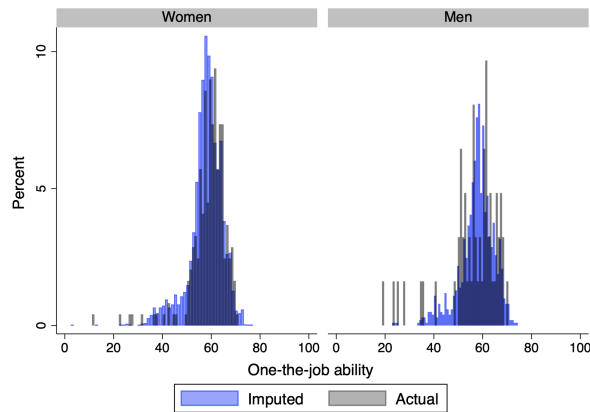
Table A.18. Effort in application completion

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Access to portal	# edits	% completed	Question 1 length	Question 2 length
Panel A: Men					
Male Photo	0.029 (0.025)	1.522 (2.125)	-0.029 (0.030)	-23.168 (55.645)	-35.240 (46.005)
Lower Share	0.008 (0.025)	4.553 (2.185)	0.029 (0.030)	32.439 (55.686)	41.648 (46.057)
Photo = Info p	0.54	0.30	0.17	0.48	0.24
N	804	687	807	807	807
Mean DV	0.85	29.01	0.55	976.9	823.5
Mean DV in PC	0.85	27.75	0.53	978	795.6
Panel B: Women					
Male Photo	-0.037 (0.011)	0.650 (1.066)	-0.039 (0.014)	-66.878 (25.893)	-58.321 (21.412)
Lower Share	-0.012 (0.011)	-0.004 (1.067)	-0.009 (0.014)	-16.895 (25.867)	-8.457 (21.384)
Photo = Info p	0.13	0.69	0.13	0.17	0.09
N	3503	3063	3513	3513	3513
Mean DV	0.87	33.01	0.56	1019.89	849.01
Mean DV in PC	0.88	31.85	0.56	1029.84	855.02
<i>Gender differences</i>					
Photo: Men=Women p	0.02	0.71	0.75	0.47	0.65
Info: Men=Women p	0.47	0.06	0.24	0.42	0.32

Robust standard errors in parentheses

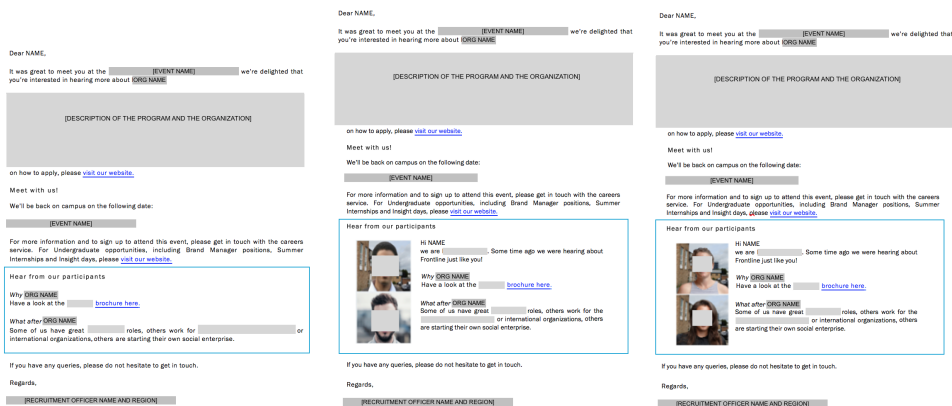
Note. The table shows OLS estimates for proxies of effort in application completion. The variable “Access to portal” is a dummy for whether the person ever accessed the application portal to edit the application form. The variable “# edits” counts how many times a candidate logged-in to make changes to the application form. “% completed” is percentage of non-blank fields in the application form. The variables “Question 1 length” and “Question 2 length” count the number of characters used in each of the two motivational questions (open-ended) contained in the application. All the regressions control for stratification variables and dummies for the week in which the candidate registered. See Table 2 for the definition of the other rows.

Figure A.14. Comparison of predicted and actual on-the-job performance



Note. The figure shows the comparison of predicted and actual on-the job performance distributions. The histograms on the LHS are for women and on the RHS for men. Ability is on a scale from 0 (min) to 100 (max). The methodology to construct this measure is described in Appendix B.3.

Figure A.15. Additional experiment in universities: treatments



Note. The figure shows the emails used in the additional experiment in universities described in Appendix D. One treatment showed two photographs of male workers (email in the center of the figure), and a second treatment two photographs of female workers (last email on the right of the figure).

Table A.19. Additional experiment in universities: balance and summary statistics

	(1)	(2)	(3)	(4)	(5)	(6)
	Overall		Joint test		Pairwise tests	
	<i>N</i>	<i>mean</i>	<i>sd</i>	<i>F stat</i>	<i>p-value</i>	<i>min p value</i>
Male	2877	0.22	0.42	1.49	0.23	0.09
Last year	2500	0.57	0.49	0.07	0.93	0.72
Graduates	2500	0.11	0.31	0.18	0.84	0.59
First/second year	2500	0.32	0.47	0.02	0.98	0.85
Science or business	2334	0.21	0.41	1.22	0.3	0.16
Heard about the job/organization	2334	0.29	0.45	1.09	0.34	0.19
- on campus	1221	0.21	0.41	1.21	0.3	0.18
- in news/ads	1221	0.55	0.5	0.91	0.4	0.18
- from friends	1221	0.07	0.26	0.15	0.86	0.60
- online	1221	0.17	0.37	0.5	0.61	0.35

Note. The table shows summary statistics and balance checks for the additional experiment in universities described in Appendix D. “First/second/last year” are indicators for the year of enrolment in university. “Science or business” is an indicator for studying a scientific or economics/business subject. “Heard about the job” is equal to one if the person heard of the organization before attending the event. Columns 4 and 5 report the F-statistic and p-value from a joint test of the significance of the treatment dummies in explaining each row variable. The last two Columns report the minimum p-value from t-tests between pairs of treatments.

Table A.20. Additional experiment in universities: results

VARIABLES	(1)	(2)	(3)	(4)
	Clicked to website		Applied	
Sample:	Men	Women	Men	Women
Female Photo	-0.019	0.007	-0.019	0.019
	(0.030)	(0.017)	(0.018)	(0.019)
Male Photo	0.076**	0.021	0.000	0.012
	(0.030)	(0.016)	(0.018)	(0.019)
Male=Female Photo p	0.0	0.39	0.27	0.71
Observations	209	286	475	1,612
Mean DV in PC	0.06	0.09	0.04	0.12
<i>Controls</i> : University-Event FE	✓	✓	✓	✓

Standard errors in parentheses

Note. The table shows the main results of the additional experiment in universities described in Appendix D. The dependent variable in Columns (1) and (2) is the share of participants in each event-treatment group who clicked on a link (only new users), and in Columns (3) and (4) is a dummy for applying for the job. The regression in Columns (1) and (2) is at the university-event-gender-treatment level, while in Columns (3) and (4) is at the individual level. The omitted category is the group receiving emails with no workers’ photographs. “Female Photo” and “Male Photo” are indicator variables for the two experimental treatments. The regression includes university-event fixed effects and analytical weights by treatment group size. The last two Columns limit the sample to students who are eligible to apply (in their last year or graduates). “Male=Female Photo p” is the p-values of a test of equality of the coefficients on the male and female photograph.

B Methodology for the creation of indexes

B.1 Index of desirable skills

Figures A.4, A.5a and A.6 use an “index of desirable skills” as a measure of the quality of the pool of applicants. The index is computed as the mean of the following standardized variables: receiving a first grade in university, being from a top UK university, being in full time employment, having graduated before 2017, studying a common subject for social work, having volunteered frequently in the past, level of cognitive skills and score in English pre-university tests. Each component is standardized in the overall sample of applicants, and then the average of all the components is re-standardized again.

Most of the variables used in the index are reported by the candidates in their registration or application form. Three variables are constructed: being from a top UK university, cognitive skills and having studied a common subject for social workers. I define top UK universities as those belonging to the Russell Group (which is the same classification used by the partner organization. See: <http://russellgroup.ac.uk>). To measure cognitive skills, I use the information reported by candidates in their application form, where they were asked to describe recent jobs and employers. With the help of three RAs, I code the most recent job into standardized SOC4 categories (coding was blind to treatment status). I then follow the methodology of Acemoglu and Autor (2011), and use O*Net data (National Center for O*NET Development, 2019) to classify the tasks of each occupation into cognitive or manual, routine or non-routine. For each candidate, the cognitive skills score corresponds to the score for their most recent job. To define which university subjects are common for social workers, I first looked at the knowledge areas listed for social work by O*Net and then double checked the classification with the partner organization.

I chose the components of the index to be variables that are positively and significantly correlated with receiving a job offer, which indicates that they are “desirable” skills from the employer’s perspective (see Table A.17). In addition, I also ran a Lasso regression of job offer on many different demographic and educational observables. The components included in the index belong to the intersection of the variables which are significant in the OLS regression of Table A.17 and are also selected by the Lasso (except for stratification variables, which I control for in Figure A.4 in the paper). The only exception is studying a common subject for social work, which is not selected by the Lasso. I decided to keep this last variable into the index given the importance that the employer gives to it during screening, but results are similar when excluding this additional component from the index. The lasso results are the same when using the unconditional offer variable, or the variable conditional on applying (which is missing for those who did not apply).

B.2 Predicted outside option: wage level and dispersion

I compute the individual expected hourly wage in the UK as a measure of the outside option level. This measure is used in Figure A.7 and in Table A.13 of the paper. Using the Labour Force Survey (LFS) quarterly data between January 2017 and December 2018 (Office for National

[Statistics, 2023](#)), I estimate a Mincerian regression of the log-hourly wage on a set of observables which are available both in the LFS and my experimental dataset. I then impute the coefficients of the Mincerian regression to my data to predict an individual-level expected wage.

I limit the LFS sample to men and women between 16 and 64 years old and, to match the eligibility criteria of my organization, I include only people who have at least a bachelor degree or, if students, who are currently studying towards a bachelor degree or higher university title. The variable (HOURPAY) for the hourly pay is truncated between 0 and 99 and is computed for all respondents who are employees and those on a government scheme. I estimate the regression of the log-hourly wage on the following set of dummies: university subject (16 categories), age, age squared, British nationality, gender, marital status, non-white ethnicity, first grade in university. The omitted category are non-married white women who studied Arts.

The measure of wage dispersion (used in Table 4) is also computed using the LFS data. I first compute the 75/25 interquartile range of wages in each industry in the UK. There are a total of 21 industries in the Labor Force Survey, each corresponding to a unique SIC2007 code (indicated by the variable INDS07M). For a candidate from field of study s , I then compute a weighted average of the interquartile ranges across all industries, where the weights are given by the share of s graduates working in each industry. The intuition is that the weights represent the probability for a candidate from field of study s to enter that particular industry, and thus to face that industry-specific wage dispersion in his/her outside option. I coded the field of study of participants in my experiment using the same categories of the LFS data through a fuzzy-matching using data from HESA ([Higher Education Statistics Agency, 2017b, 2018](#)).

B.3 Predicted performance on the job

I construct the predicted on-the-job performance score based on observables through a linear truncated regression in the pure control group. The truncated regression uses the following variables: ranking and completion rate of the candidate’s university, subject studied, obtaining a first grade, whether the grade is confirmed, having graduated, age, age squared and full-time employment. Data on universities are from the [Complete University Guide \(2019\)](#). This is the most comprehensive set of baseline variables which is also available in my survey data. Figure A.14 shows the distribution of predicted performance against actual scores. To consider this predicted value as a valid measure of the counterfactual performance that a person would achieve in the job, the assumption is that the observables used in the prediction affect job performance independently of being hired and treatment.

B.4 Index of occupational gender segregation

I compute the Duncan index of occupational segregation ([Duncan and Duncan, 1955](#)) using microdata on the occupational structure by gender from a 10% random sample of the 2011 UK Census ([Office for National Statistics, 2015](#)). The data contain the distribution of workers across 362 detailed SOC4 occupational categories at the Medium Layer Super Output Areas (MSOA) level. There are 7201 MSOAs in the data: the median MSOA comprises 188 7-digits postcodes, with a minimum of 89 to a maximum of 1033. The index is computed as: $\frac{1}{2} \sum_{i=1}^N \left| \frac{m_i}{M} - \frac{f_i}{F} \right|$,

where m_i and f_i are the male and female population, respectively, in occupation i and M and F are the total working population in the local labor market.

Using a bridge between MSOAs and 7-digit postcodes, I merged the index with my experimental data through the subjects' secondary school postcode and, when missing (for 62% of the sample), home postcode. For half of the sample with missing school information, home postcodes capture the place where they grew up. For the other half, it is the current domicile. In case of no match between MSOA and 7-digit postcodes, I use 3-digit postcodes and impute the modal MSOA Duncan Index to the respondent. The distribution of the Duncan index in my experimental sample is similar to the UK one: the former has a mean index of 0.56 (SD=0.08) and the latter of 0.58 (SD=0.07).

C Out-of-trial surveys: implementation

The following paragraphs describe the implementation and recruitment procedures of the surveys used in the paper. I collected surveys 1, 2 and 3 between July and December 2018, while surveys 4 and 5 were collected in January 2023 and survey 6 in October 2020. Pre-analysis plans for surveys 2 and 3 are available in the AEA RCT Registry (Delfino, 2021b). Surveys 2, 3 and 4 are primarily discussed in Section V. The data from survey 1 are used only in Table A.1 and the data from survey 5 are only used in Figure A.2c. The evidence from survey 6 is discussed in Section IV.C.

Survey 1 - Amazon Turk. This survey tests for differences between photographs on a sample of 161 Mechanical Turk workers, in order to check whether images differed in some dimensions other than gender. Respondents i) are located anywhere in the world, ii) have not participated in any of the researchers' previous experiments and iii) hold the "Master" qualification. The survey took an average of 2 minutes and was rewarded with 20 cents. Another survey - similar to 2 and 3 - was planned on Amazon Turk, but was not pursued.

Survey 2 - 2018 Applicants. One year after the field experiment, I worked with the partner organization to invite new candidates for the job to participate in my online survey. Invitations were sent to approximately 4500 people over two days. The sample comprises candidates at different stages of the selection process who registered their interest between September and November 2018. As incentive to participate, I compensated the first 300 respondents with \$5, which they could keep for themselves or donate to a charity. All the participants were also automatically enrolled into a raffle for a \$150 Amazon voucher. A total of 303 people fully completed the survey, which corresponds to a response rate of around 7%. While men's share is equal to the population - less than 20% - their number is too small to do analyses by gender in this sample.

Survey 3 - Prolific. Respondents are Prolific users who i) have not participated in any of the researchers' previous surveys, ii) are British, iii) have an approval rate above 75 %, iv) are

between 18 and 64 years old and v) have at least a bachelor degree. The final sample includes 130 women and 131 men. I collected answers in different waves to match the composition of the participants in the field experiment on the following observable criteria: gender, ethnicity, employment/student status, university subject, job sector. The participation fee was \$1.50.

Both surveys 2 and 3 were conducted with a between-subject design and comprised almost the same set of questions. I first showed respondents a photograph and asked short questions about the portrayed worker. Then participants looked at one intervention email (using the same photograph) for at least 30 seconds. After mandatory understanding checks, I elicited beliefs on a variety of dimensions about the advertised job, its applicants and workers. Randomization was stratified by gender and ethnicity, and I ensured that the respondent's ethnicity matched that of the worker depicted in the photograph (as in the main field experiment). In survey 2, 61 women were also assigned to a control group which saw no experimental email. Average completion time was 15 minutes.

Survey 4 - Prolific. I surveyed 65 men and 23 women recruited on the platform Prolific, with sampling following criteria i) to iv) from survey 3. In a between-subject design, I showed each respondent one of the intervention emails and asked open-ended questions about the interpretation of the information treatment. The participation fee was \$1.20 for an average completion time of 7 minutes.

Survey 5 - Prolific. I surveyed 201 workers to provide evidence on men's and women's expectations on returns to talent in female- versus male-dominated jobs. Sampling was balanced by gender (N Male= 99, N Women = 102) and stratified by ethnicity (76% white and 24% non-white) and age (75% between 19 and 34, 15% between 35 and 44, 5% between 45 and 54 and 5% above 55). I oversampled young respondents to reflect the demographic composition of my RCT sample and, more broadly, UK job-seekers. The participation fee was \$1.20 for an average completion time of 7 minutes.

Survey 6 - Personnel of the partner organization. I sent a survey to the management team of the partner organization to gauge their opinion regarding the trade-off between performance and retention in the program (N=31). Half of the respondents are supervisors of social workers in local communities and half belong to recruitment functions. 55% of the respondents have been working in the organization for less than two years and 45% of them have managerial responsibilities. I used two vignette-type questions, shown in Figure [A.10](#). The statistics used in the two scenarios come from data on women from treatment groups "Woman Photo + Higher Share" and "Male Photo + Lower Share".

Surveys 2 and 3: questions and definition of main variables

Variable	Survey Question	Measurement
Photographs-related beliefs		
% female applicants	Consider 100 people who apply after seeing the email ad you have just seen. Of these people, how many do you think are women?	
Employer wants own gender	Think about the moment in which you first saw the email ad... To what extent did you think that the organization wants to attract people of your gender?	5-points scale from "Not at all" to "A great deal". Analysis: dummy variable for top three choices.
Discrimination by customers	By looking at this ad, to what extent do you agree with the following statement? "Customers discriminate workers (by race or gender) in this job."	6-points scale from "Strongly Agree" to "Strongly Disagree". Analysis: dummy variable for top three choices.
Job desirable for men	By looking at this ad, to what extent do you agree with the following statement? "The job is desirable for a man."	As above
Job desirable for women	By looking at this ad, to what extent do you agree with the following statement? "The job is desirable for a woman."	As above
Information-related beliefs		
% of high-perf applicants	Consider 100 [women/men] who apply for this job after seeing the email ad that you have just seen. Of these 100 [women/men], how many do you think that have the potential to get commendable or excellent feedback on the job?	
% of high-performing workers	Now that you have seen again the email ad... Indicate the proportion of [women/men] that you think are successful on-the-job. Interpret "success" as people who got commendable or excellent feedback on the job.	Scale from 0-10% to 91-100% (10 points). Analysis: continuous variable using each option's middle points.
Self-reported ability	How do you expect a person with your skills and experience to perform when interacting with families in need?	Scale 1 (lowest) to 10 (highest)
Own exp ranking for the job	Consider 100 [women/men] who are applying for this job. Based on the ad you just viewed, how would you rank yourself for the job among them?	Scale 1 (best) to 100 (worst). Analysis: reverse coded.
Employer's exp ranking for the job	Consider 100 [women/men] who are applying for this job. Based on the ad you just viewed, how do you think the employer would rank you for the job among them?	Scale 1 (best) to 100 (worst). Analysis: reverse coded.
Job difficulty	By looking at this ad, to what extent do you think that the job is difficult?	Scale 0 (very easy) to 100 (very difficult)
Quality standards	By looking at this ad, to what extent do you think that the organization requires high quality standards by its employees in the relationship with customers?	Scale 0 (very low quality) to 100 (very high quality)
Promotion difficulty	By looking at this ad, how easily do you think that people get promoted in this job?	Scale 0 (very easy) to 100 (very hard)
Application effort	By looking at this ad, how much effort do you think that the application process requires?	Scale 0 (very low) to 100 (very high)
Number of applicants	Consider 100 people who are considering whether or not to apply for this job. How many do you think will apply?	
Other beliefs		
Wage level	By looking at this ad, do you think that the job has a high or low wage?	Scale 0 (low wage) to 100 (high wage)
Job high social status	By looking at this ad, to what extent do you agree with the following statement? "The job has a high social status."	6-points scale from "Strongly Agree" to "Strongly Disagree". Analysis: dummy variable for top three choices.
Felt fooled by ad	Think about the moment in which you first saw the email ad... To what extent did you worry that you were being fooled?	5-points scale from "Not at all" to "A great deal". Analysis: dummy variable for top three choices.

D External validity of the photographs effect: a complementary field experiment

During 2017 I conducted another field experiment with the same partner organization in order to try and increase the pool of candidates interested in the job. This second experiment was pre-registered together with the main field experiment presented in the paper and focused on testing whether gendered photographs could encourage students to get informed about the job and eventually apply for it. The results of this second experiment can thus provide evidence on whether the null effect of male photographs on men’s applications generalizes in a sample of young men who are not yet selected on interest in the job.

Setting: In the fall of 2017, the partner organization visited 52 universities across the UK to promote its program and encourage applications in different events (e.g., job fairs, workshops). On average, each university was visited three times (max 6). Each university is assigned to a Recruitment Officer (RO) in charge of organizing and conducting the events, collecting email addresses of participants and sending a follow-up email with further information about the job.

Experiment: People who took part to events and left their contacts in a mailing list were randomly assigned to three groups, which differed in the format of the follow-up email received. The text in these three emails was exactly the same, but they might show i) no photographs, ii) photographs of female workers, ii) photographs of male workers. The three email templates are shown in Figure A.15. Assignment to treatment was stratified by university, event and gender.

Outcomes and balance: Each email contains links to the organization’s website which are trackable at the stratum-treatment level. This allows me to know the number of participants of gender g in event e in university u that clicked on any email link, whether they are first time users and some metrics of their online behavior, for each treatment group. Each event had an average number of 30 sign-ups, for a total of 2877 unique participants. Table A.19 presents summary statistics of the experimental sample and balance checks. The main outcome for application is whether people apply to the job offered by the partner organization.

Specification: For a group of gender g , within event e and university u I estimate:

$$y_{geu} = c + \beta_1 MPhoto_{geu} + \beta_2 WPhoto_{geu} + \delta_{eu} + \epsilon_{geu}$$

where δ_{eu} are university-event fixed effects and $MPhoto_{geu}$ ($WPhoto_{geu}$) is a dummy for the male (female) photographs. Estimates are weighted by treatment group size.

Results: Table A.20 shows that men are twice as likely to access the organization’s website when they receive an email with male photographs as compared to both the control group and female photographs. However, male photographs don’t increase men’s applications. This increases the external validity of the null effect of the “Male Photo” in my main experiment.

E Information interpretation: evidence from a survey with open-ended questions

I ran an additional out-of-trial survey on Prolific in January 2023 (described as “Survey 4” in Appendix C) to explore the interpretation of the information treatment through open-ended questions. Here are the questions I asked:

- 1 If you were a potential applicant, how would you interpret the information given in the ad? That is: [66% / 89%] of people got high evaluations in a previous cohort.
Now keep thinking that you are a potential applicant and you see that [66% / 89%] of people got high evaluations in a previous cohort.
- 2 What types of applicants do you think the employer is trying to attract?
- 3 What would you think about the level of competition in the job?
- 4 What would you think about the effectiveness of previous workers and whether there is room for improving it?

A qualitative exploration of the answers to the first two questions shows the following patterns:

- Respondents interpret both statistics positively. In either treatment, most of them say that previous workers achieved “high levels” or “high standards”. For instance, some answers say that the information is “Pretty positive, relatively high figure for high evaluations” (in the 66% info) and that the ad shows “That the person in the ad was a qualified candidate and was able to do the job advertised to a high standard” (in the 89% info).
- For both statistics, most respondents think that the employer is trying to attract high-skilled, hard-working applicants. For instance, some respondents say “I think they are trying to attract hard working individuals who want to excel in their field of work” (in the 89% info) or that the employer is looking for people who are “Excellent in dealing with people or families” (in the 66% info);
- A few people see the 66% information concerning in terms of the employer’s standards. For instance, one respondent says: “I am unsure whether it is trying to suggest it’s hard to get high mark”. Another difference between the 66% and 89% information is in the extent to which people can change outcomes for the employer. For instance, one respondent in the 66% treatment mentions “I would think that the post does not currently have a high number of talented applicants and that given 66% were given high evaluations that the employer would be looking for a more talented worker.” Another mentions “66% is a quite high number and as an applicant it would entice me to apply. It would also encourage to not only meet that target but try to exceed it”. In the 89% treatment instead there are a few people who say doubt the truthfulness of information, along the lines of: “there is a lack of skilled people wanting to do this job, and every score is exaggerated to keep everyone who applied for this job happy.” Similarly, another person in the 89% treatment mentions that the employer is targeting “It is saying that almost 9 in 10 people passed, and so the chance of the applicant getting a high evaluation is also high”.

Regarding the two questions explicitly asking about competition and effectiveness:

- The level of competition is deemed to be high for both treatments, with only 10 respondents reporting that they believe competition is low.
- Regarding effectiveness, workers in the 89% treatment are considered more effective and reaching “impressive” results. While most respondents say that there is always room for improvements (even if small), most people consider the potential improvements to be larger in the 66% treatment. For instance, one person said “66% is a strong number for highly evaluated candidates. Although there is still room for improvement (in regards to increasing the % of high evaluations) it still sets a high bar. There is always room for improvement and new candidates can be the ones to make that change from past workers.”

F Additional alternative treatment interpretations

In this section I discuss additional alternative interpretations of the two treatments, considered overall less relevant compared to those discussed in section V of the paper.

Risk aversion. A higher variance in success among past workers (66%) could increase the perceived riskiness of performance in the job, leading to different reactions by gender (Eckel and Grossman, 2008). However, such higher variance may imply lower risk if it signals that there is a stronger link between talent and evaluations, as suggested in Section V.B. Similarly, the “Higher Share” could be associated with higher uncertainty if people think that success is determined by other (unclear) factors rather than ability. Even allowing the “Lower Share” to be associated with higher risk, we should expect women to apply less in this condition on average.

On-the-job dating market. Experimental photographs may be a signal of dating opportunities. If so, we expect heterosexual (non-heterosexual) men to apply more when seen a female (male) photograph and weaker effects on married people. Delfino (2021a) shows that marital status does not affect the reaction to photographs. Qualitatively, non-heterosexual men and women both react positively to the “Male Photo”, implying that the interpretation of photographs is not related to dating.

Overconfidence. Do men and women react differently in the “Lower Share” group because of differences in overconfidence? The main concern is that more men apply because they are overconfident and wrongly believe they will perform better than past workers. Results on actual job performance contrast with this view. Moreover, men in my sample tend to be less overconfident than women in job-specific skills, which is aligned with evidence that the gender gap in overconfidence is mitigated and even reverses in female domains (Coffman et al., 2019).

Moreover, Delfino (2021a) shows that applications increase the most among men with low confidence in their beliefs of people’s performance in social work. Thus the effects seem driven by the least confident men, as long as confidence about others’ performance is correlated with confidence in one’s own abilities.

References

- Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier.
- Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.
- Coffman, K. B., Collis, M., and Kulkarni, L. (2019). Stereotypes and belief updating. *HBS Working Paper*, (19-068).
- Complete University Guide (2019). University league tables. <https://www.thecompleteuniversityguide.co.uk/league-tables/rankings>.
- Delfino, A. (2021a). Breaking gender barriers: Experimental evidence on men in pink-collar jobs. *IZA Discussion Paper 14083*.
- Delfino, A. (2021b). Social identity and career choices. AEA RCT Registry [Publisher]. doi: <https://doi.org/10.1257/rct.2351-5.1999999999999999>.
- Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American sociological review*, 20(2):210–217.
- Eckel, C. C. and Grossman, P. J. (2008). Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results*, 1:1061–1073.
- Flood, S., King, M., Rodgers, R., Ruggles, S., Warren, R., Backman, D., Chen, A., Cooper, G., Richards, S., Schouweiler, M., and Westberry, M. (2023). IPUMS CPS: Version 11.0 [dataset]. doi: <https://doi.org/10.18128/D030.V11.0>.
- Higher Education Statistics Agency (2017a). He student enrolments by subject area, principal subject, level of study, mode of study, domicile, sex and academic year. <https://www.hesa.ac.uk/data-and-analysis/students/table-9>.
- Higher Education Statistics Agency (2017b). Unistats record 2017/18 (version 1.15). <https://www.hesa.ac.uk/collection/c17061>.
- Higher Education Statistics Agency (2018). Jacs 3.0: Principal subject codes. <https://www.hesa.ac.uk/support/documentation/jacs/jacs3-principal>.
- List, J. A., Shaikh, A. M., and Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4):773–793.
- National Center for O*NET Development (2019). O*Net 23.3 database. https://www.onetcenter.org/db_releases.html.
- Office for National Statistics (2015). 2011 census: CT0347 sex by occupation (4 digits) national to MSOA. The National Archives [Publisher]. <https://webarchive.nationalarchives.gov.uk/ukgwa/20160110200035>.
- Office for National Statistics (2020). Earnings and hours worked, region by occupation by four-digit SOC: ASHE table 15.
- Office for National Statistics (2023). Labour force survey. [data series]. 9th release. UK Data Service [distributor]. doi: <http://doi.org/10.5255/UKDA-Series-2000026>.
- Projections Central (2016). Long-term occupational projections (2016-2026). Projections Managing Partnership [Publisher]. <https://projectionscentral.org/Projections/LongTerm>.
- US Bureau of Labor Statistics (2017). Employment status of the civilian noninstitutional population in states by sex, race, hispanic or latino ethnicity, marital status, and detailed age. <https://www.bls.gov/lau/ex14tables.htm>.
- US Bureau of Labor Statistics (2018a). Occupational employment and wage statistics (national). <https://www.bls.gov/oes/tables.htm>.
- US Bureau of Labor Statistics (2018b). Women in the labor force: a databook. Technical report.
- US Bureau of Labor Statistics (2019). Employment projections. <https://www.bls.gov/emp/>.